

UNIVERSIDADE FEDERAL DO PARANÁ

WILLIAM RODRIGO JOANICO

**ANÁLISE AUTOMATIZADA EM GÉIS DE ELETROFORESE UNIDIMENSIONAIS
UTILIZANDO TÉCNICAS DE PROCESSAMENTO DE IMAGENS - AAGEU**

CURITIBA

2012

WILLIAM RODRIGO JOANICO

**ANÁLISE AUTOMATIZADA EM GÉIS DE ELETROFORESE UNIDIMENSIONAIS
UTILIZANDO TÉCNICAS DE PROCESSAMENTO DE IMAGENS - AAGEU**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Bioinformática, Setor de Educação Profissional e Tecnológica, da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de Mestre em Bioinformática.

Orientador:
Iris Hass, Dra.
Co-orientador:
Lucas Ferrari de Oliveira, Dr.

**CURITIBA
2012**

*Dedico esse trabalho à minha família,
que sempre me incentivou e deu apoio à minha
formação; e a Deus, pela vida: a força propulsora
das minhas atividades e ao mesmo tempo objeto
do meu estudo.*

AGRADECIMENTOS

Registro meus agradecimentos a todos os que compartilharam o trilhar de mais esse caminho percorrido, contribuindo, direta e indiretamente, para que eu realizasse esta pesquisa, auxiliando-me e dando-me forças nos momentos em que mais precisei.

Agradeço, especialmente, à minha família, pelo apoio para que eu concretizasse essa pesquisa: minha mãe Rosalina, meu pai Carlos, minhas irmãs Danielle e Rafaella, minha tia Ivone e ao meu grande amigo\irmão William que foram incansáveis e estiveram ao meu lado, entendendo-me nos momentos de ausência, dando-me apoio e carinho.

À professora Doutora Iris Hass, minha orientadora, que possibilitou-me “aprendizagens únicas”, por meio do grande incentivo e orientação que me foram concedidos durante essa jornada.

Agradeço, imensamente, ao meu co-orientador Prof. Dr. Lucas Ferrari de Oliveira, por eu ter chegado até o final, revendo alguns pré-conceitos, refletindo sobre alguns pontos de vista e principalmente por ser tão paciente, me propiciou um grande aprendizado, pela assistência, apoio e tempo dedicado durante este trabalho. Para mim foi um grande prazer poder trabalhar ao seu lado.

Ao aluno de Iniciação científica, Alysson Narloch, pelo acompanhamento durante os experimentos, dando suporte para o funcionamento do sistema e desenvolvimento de ferramentas que foram essenciais para o êxito desse trabalho.

Aos colegas e professores do mestrado, por tudo o que com eles aprendi e por partilharem a construção do meu estudo. Em especial, aos meus colegas de laboratório Sérgio, Lucas, Juliana, Vanelly, Leviston, Paula, Jessé, Danhylo, Katia, Barbara, Rafael e Rosa, valeram os momentos de conversas, discussões e distrações.

À Kelly Rafaela, amiga e colega de laboratório, que sempre estava presente, me auxiliando, dando dicas e sugestões, pelos momentos difíceis e de indecisões, e também por boa parte da alegria que era compartilhada no ambiente de pesquisa.

À professora Dr. Ana Carolina Pilatti que me confiou a carta de recomendação a este programa e pelo auxílio na aquisição de conhecimentos durante minha graduação.

Ao professor e amigo Sandro Germano, que durante o mestrado me incentivou aos estudos e me ajudou em momentos difíceis.

À Suzana pela amizade, compreensão, ajuda e pela vontade de sempre ajudar.

Ao programa de Pós Graduação em Bioinformática pela oportunidade.

À Capes pelo auxílio financeiro.

Enfim, a todos aqueles que direta ou indiretamente contribuíram para que este trabalho chegasse ao fim, os meus mais profundos agradecimentos.

.

RESUMO

Este trabalho apresenta uma abordagem computacional automatizada para realizar a detecção de estruturas pertencentes às imagens de géis de eletroforese unidimensionais, como colunas e marcadores. Foi desenvolvido um algoritmo em linguagem C, com auxílio da biblioteca OpenCV para processar as imagens. As imagens foram submetidas a uma fase de pré-processamento com recortes, alinhamentos e aplicação de um filtro bi-lateral. A metodologia abordada para detectar o marcador fez uso da imagem segmentada e da informação contida em cada coluna. No caso da obtenção das colunas a imagem binarizada passou por um processo de abertura morfológica e em seguida cada coluna foi representada por um vetor e baseado em seu conteúdo delimitadas as colunas. Os resultados demonstram que a metodologia utilizada é válida, e que através de técnicas simples é possível resolver o problema, sem a necessidade de se desenvolver ou utilizar técnicas mais complexas ou altamente custosas computacionalmente. A técnica obteve 73% de acertos na identificação do marcador, com base nos testes realizados e as dificuldades encontradas estavam relacionadas com o contraste e a presença de ruídos e manchas em algumas imagens. No caso do algoritmo para a identificação das colunas, a técnica alcançou 82% de acertos e foi possível notar que o algoritmo conseguiu identificar todas ou pelo menos a maior parte das colunas que estavam presentes nas imagens, e esse sucesso foi devido aos poços estarem alinhados e as imagens possuírem boa qualidade, assim evidenciando suas informações.

Palavras-chave: géis de eletroforese unidimensionais, marcadores, colunas, linguagem C.

ABSTRACT

This paper presents a computational approach to perform automated detection of structures belonging to the images of one-dimensional gels electrophoresis like pillars and markers. We developed an algorithm in C language with the aid of the OpenCV library to process the images. The images were submitted to a phase of preprocessing with cutouts, alignments and applying a filter bi-lateral. The methodology discussed to detect the marker made use of the segmented image and the information contained in each column. In the case of columns to obtain the binarized image has undergone a process of morphological opening and then each column is represented by a vector-based and content-defined columns. The results show that the methodology is valid, and that simple techniques can solve the problem without the need to develop or use more complex techniques or highly expensive computationally. The technique achieved 73% accuracy in identifying marker, based on tests and the difficulties encountered were related to the contrast and the presence of noise and spots on some images. Under Algorithm for the identification of columns, the technique reached 82% of correct answers and was possible to notice that the algorithm able to identify all or at least most columns that were present in images, and that success was due to the wells being aligned and the images were good quality, so showing your information

Keywords: One-dimensional electrophoresis gels, markers, columns, C language.

LISTA DE FIGURAS

FIGURA 1.1 - EXEMPLO DE UM GEL RESULTANTE DE UMA ELETROFORESE	13
FIGURA 2.1 - TÉCNICA DE ELETROFORESE	19
FIGURA 2.2 - APARELHOS UTILIZADOS NA TÉCNICA DE ELETROFORESE VERTICAL EM GEL DE POLIACRILAMIDA.	20
FIGURA 2.3 – BANDAS E COLUNAS	20
FIGURA 2.4 - EXEMPLO DE IMAGEM UTILIZADA NESSE TRABALHO	21
FIGURA 2.5 - EXEMPLO DA CONVENÇÃO DOS EIXOS PARA REPRESENTAÇÃO DE IMAGENS DIGITAIS	27
FIGURA 2.6 - RESOLUÇÃO X PROFUNDIDADE	28
FIGURA 2.7 - ETAPAS NO PROCESSAMENTO DE IMAGENS	29
FIGURA 2.8 - FUNÇÃO DO HISTOGRAMA DE UMA IMAGEM $F(X, Y)$ DE DIMENSÃO $N \times M$	31
FIGURA 2.9 - EXEMPLO DE HISTOGRAMAS	32
FIGURA 2.10 – DEFINIÇÃO DO FILTRO GAUSSIANO (GB)	33
FIGURA 2.11 – DISTRIBUIÇÃO GAUSSIANA COM MÉDIA ZERO E DESVIO UM, $\text{SIGMA} = 1$, NA EQUAÇÃO DA FUNÇÃO GAUSSIANA	34
FIGURA 2.12 - RESULTADO COM O USO DO FILTRO GAUSSIANO COM DIFERENTES VALORES PARA O SIGMA.	35
FIGURA 2.13 – DEFINIÇÃO DO FILTRO BILATERAL (BF).	36
FIGURA 2.14 – WP COMO SENDO UM ELEMENTO DE NORMALIZAÇÃO	37
FIGURA 2.15 – EFEITO DOS PARÂMETROS ESPACIAIS (SIGMA_S) E DO ESPAÇO DE COR (SIGMA_R) E UMA CONFRONTAÇÃO COM O FILTRO GAUSSIANO	38
FIGURA 2.16 - EXEMPLO DE SEGMENTAÇÃO	41
FIGURA 2.17 - EXEMPLOS DE ELEMENTOS ESTRUTURANTES:	43
FIGURA 2.18 - EROÇÃO DE UM CONJUNTO X PELO ELEMENTO ESTRUTURANTE B	43
FIGURA 2.19 - EXEMPLO DE EROÇÃO	44
FIGURA 2.20 - A DILATAÇÃO DE UM CONJUNTO X PELO ELEMENTO ESTRUTURANTE B	45
FIGURA 2.21 - EXEMPLO DE DILATAÇÃO	45
FIGURA 2.22 - A ABERTURA DE UM CONJUNTO X PELO ELEMENTO	46

FIGURA 2.23 – RESULTADO ABERTURA BINÁRIA.	47
FIGURA 2.24 - FECHAMENTO DE UM CONJUNTO X PELO ELEMENTO ESTRUTURANTE B.	47
FIGURA 2.25 –RESULTADO FECHAMENTO BINÁRIO	48
FIGURA 3.1 - ALGUMAS IMAGENS DE GÉIS DE ELETROFORESE UTILIZADAS NESSE TRABALHO.	52
FIGURA 4.1 - IMAGEM APÓS PROCESSO DE ROTAÇÃO.....	56
FIGURA 4.2 - RESULTADO DAS ETAPAS DE PRÉ-PROCESSAMENTO.....	58
FIGURA 4.3 – ETAPA IDENTIFICAÇÃO DO MARCADOR.....	59
FIGURA 4.4 – ETAPA IDENTIFICAÇÃO DAS COLUNAS	61
FIGURA 4.5 – PROBLEMA ENCONTRADO NA IDENTIFICAÇÃO DO MARCADOR (CASO 1).	62
FIGURA 4.6 – PROBLEMA ENCONTRADO NA IDENTIFICAÇÃO DO MARCADOR (CASO 2).	63
FIGURA 4.7 - PROBLEMA ENCONTRADO NA IDENTIFICAÇÃO DO MARCADOR (CASO 3).	64
FIGURA 4.8 – RESULTADOS ACIMA DE 85% OBTIDOS NA IDENTIFICAÇÃO DAS COLUNAS	66
FIGURA 4.9 - RESULTADOS ACIMA DE 85% OBTIDOS NA IDENTIFICAÇÃO DAS COLUNAS	67
FIGURA 4.10 - RESULTADOS ACIMA DE 75% OBTIDOS NA IDENTIFICAÇÃO DAS COLUNAS	68
FIGURA 4.11 - RESULTADOS ENTRE 50% E 75% OBTIDOS NA IDENTIFICAÇÃO DAS COLUNAS	69
FIGURA 4.12 – RESULTADOS OBTIDOS NA COMPARAÇÃO DA IDENTIFICAÇÃO DO MARCADOR COM O SOFTWARE SABIA.	72
FIGURA 4.13 – RESULTADO DA COMPARAÇÃO DA IMAGEM 5.	74
FIGURA 4.14 - RESULTADO DA COMPARAÇÃO DA IMAGEM 9.	75
FIGURA 4.15 - RESULTADO DA COMPARAÇÃO DA IMAGEM 10.	76

LISTA DE SIGLAS

AAGEU	- Análise Automatizada Em Géis De Eletroforese Unidimensionais Utilizando Técnicas De Processamento De Imagens
AFLP	- Amplified Fragment Length Polymorphism
BF	- Filtro Bilateral
BIT	- Binary digit
D.P.I.	- Dots per inch
DLL	- Dynamic Linked Library
DNA	- Deoxyribonucleic Acid
GB	- Filtro Gaussiano
IO	- Input\Output
IPL	- Image Processing Library
OPENCV	- Open Source Computer Vision Library
PCR	- Reação em Cadeia da Polimerase
PI	- Ponto isoelétrico
PIC	- Conteúdo de Informação de Polimorfismo
RAPD	- Random Amplified Polymorphic DNA
RFLP	- Restriction Fragment Length Polymorphism
RNA	- Ribonucleic Acid
SABIA	- Soluções Automatizadas em Bioinformática utilizando Inteligência Artificial
SDS-PAGE	- Sodium Dodecyl Sulphate – Polyacrilamide Gel Electrophoresis
SSR	- Simple Sequence Repeats
STMS	- Sítios de Microssatélites Marcados por Sequências
UFPR	- Universidade Federal do Paraná
VNTR	- Variable Number of Tandem Repeats

SUMÁRIO

1	CAPÍTULO 1	12
1.1	Introdução.....	12
1.2	Objetivos	15
1.2.1	<i>Objetivo Geral.....</i>	<i>15</i>
1.2.2	<i>Objetivos específicos.....</i>	<i>15</i>
1.3	Justificativa.....	16
2	CAPITULO 2	18
2.1	Revisão de literatura	18
2.1.1	<i>Técnica de eletroforese de DNA em gel de poliacrilamida</i>	<i>18</i>
2.1.2	<i>Marcadores moleculares</i>	<i>23</i>
2.1.3	<i>Microsatélites.....</i>	<i>24</i>
2.2	Processamento de imagens	26
2.2.1	<i>Histograma</i>	<i>30</i>
2.3	Filtros de Suavização	32
2.3.1	<i>Filtro Gaussiano.....</i>	<i>33</i>
2.3.2	<i>Filtro Bilateral.....</i>	<i>35</i>
2.3.2.1	<i>Parâmetros.....</i>	<i>37</i>
2.4	Segmentação	39
2.5	Morfologia Matemática Binária.....	41
2.5.1	<i>Erosão</i>	<i>43</i>
2.5.2	<i>Dilatação.....</i>	<i>45</i>
2.5.3	<i>Abertura.....</i>	<i>46</i>
2.5.4	<i>Fechamento.....</i>	<i>47</i>
2.6	Linguagem C.....	48
2.7	OpenCV	50
3	CAPITULO 3	52
3.1	Materiais.....	52

3.2	Métodos.....	53
3.2.1	<i>Processamento das imagens.....</i>	<i>53</i>
3.2.2	<i>Linguagem C</i>	<i>53</i>
3.2.3	<i>Biblioteca de desenvolvimento</i>	<i>54</i>
4	CAPITULO 4	55
4.1	Resultados	55
4.1.1	<i>Processamento de cada imagem</i>	<i>55</i>
4.1.1.1	Identificação do marcador	58
4.1.1.2	Identificação das colunas	60
4.1.2	<i>Resultados com o uso do algoritmo para a identificação do marcador..</i>	<i>62</i>
4.1.3	<i>Resultados com o uso do algoritmo para a identificação das colunas...</i>	<i>64</i>
4.2	Comparação\Validação com outros softwares.....	70
5	CAPITULO 5	77
5.1	Discussão	77
5.2	Conclusões	79
	REFERÊNCIAS	81
	APÊNDICES.....	86

1 CAPÍTULO 1

1.1 Introdução

Verifica-se que não basta apenas possuir conhecimento, a forma como o conhecimento é representado é de fundamental importância a sua aceitação. Da mesma forma, não basta um sistema armazenar conhecimento, é necessário fundamentalmente que ele represente este conhecimento de maneira compreensível aos seus usuários.

Ao examinar uma amostra de DNA é possível achar padrões genéticos parecidos, estes podem dar base a inclusão de um determinado indivíduo em algum grupo que possuam particularidades iguais (AYALA, 1978).

A técnica de eletroforese é um meio pela qual se pode obter essas informações importantes para os outros estudos de genética de populações, que visa principalmente estabelecer relações de semelhança ou identificar grupos que possuam as mesmas características (LOPES, 1984).

A eletroforese consiste na separação de moléculas biológicas eletroquímicas, ou seja, com auxílio de carga elétrica e produtos químicos, essas moléculas de proteínas ou fragmentos de DNA ou RNA, são submetidas a um campo elétrico ao longo de um gel de agarose ou de poliacrilamida (LOPES, 1984).

Utilizando o DNA como exemplo, de uma forma simples, uma dada molécula é quebrada em diversos fragmentos com auxílio de enzimas específicas. Estes fragmentos são depositados no gel, onde um campo elétrico é aplicado (LOPES, 1984). Todos os fragmentos possuem massa molecular e carga elétrica diferentes, fazendo com que fiquem em posições distintas ao percorrer o gel. As moléculas com maior peso ou maior carga se movimentam mais lentamente em relação às moléculas menores. Após um determinado tempo, esse processo é paralisado e o gel passa por um procedimento de coloração, fazendo com que seja possível observar em qual posição os fragmentos migraram (MARTINEZ 2008).

Na Figura 1.1, mostra um exemplo de um gel resultante de eletroforese. As galerias verticais em destaque na Figura 1.1, são conhecidas como colunas. As

bandas são as moléculas contidas em cada coluna e as posições assumidas ao cessar a corrida eletroforética.

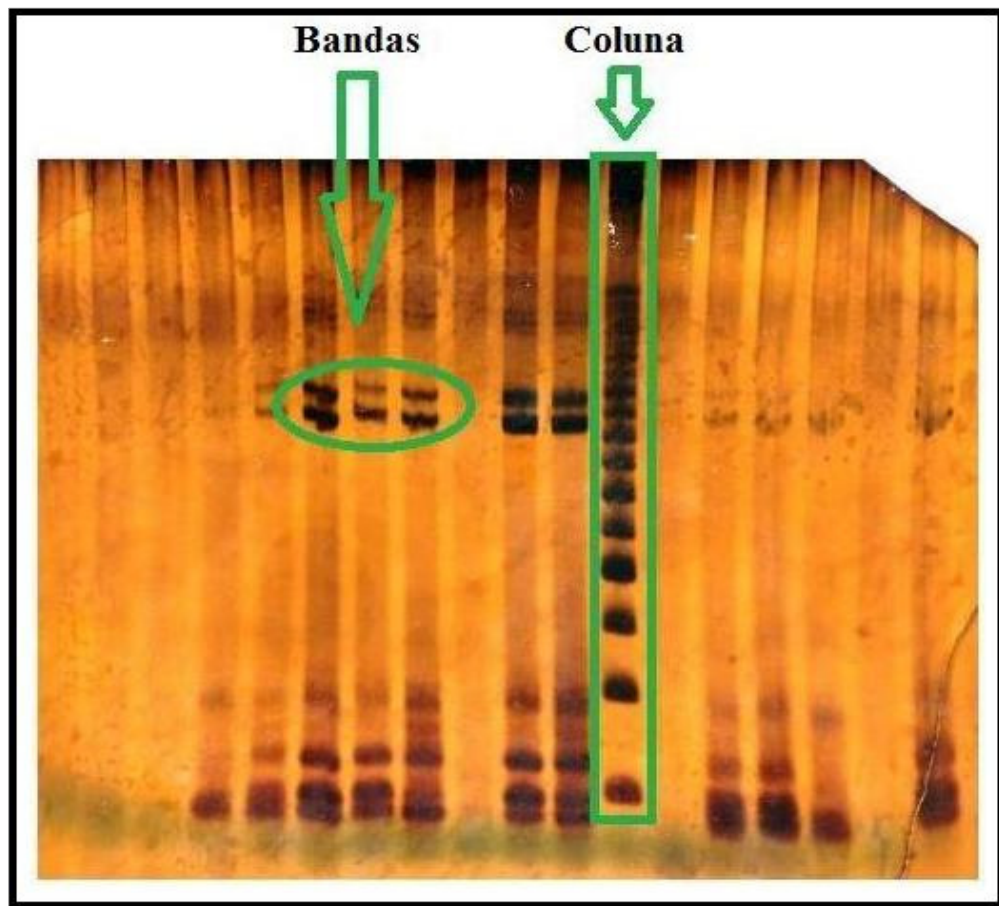


FIGURA 1.1 - Exemplo de um gel resultante de uma eletroforese

FONTE: O Autor (2011).

Com o avanço na quantidade de dados gerados através de processos informatizados que abrangem todas as áreas da biologia, a bioinformática vem na busca de soluções para enfrentar essas dificuldades que acometem o século XXI (SCHMIDT, 2010).

A pesquisa envolvendo análise de géis de eletroforese se enquadra perfeitamente neste contexto. Comumente são encontradas ferramentas para realizar esse tipo de análise, mas a sua utilização e a forma como a informação é mostrada ao pesquisador ainda necessitam ser melhoradas, para se tornarem realmente acessíveis ao cotidiano do pesquisador, já que a interação homem-dado é

uma realidade. Nenhum processo automatizado trará 100% de acertos e se for realizado uma comparação de uma determinada imagem fazendo uso de ferramentas distintas, os resultados sempre serão diferentes, pois cada um utiliza uma metodologia diferente para conseguir chegar a um resultado aceitável (KONAREV, 1981).

A análise de géis unidimensionais de eletroforese vem crescendo constantemente e com isso necessitam de soluções informatizadas mais robustas e automatizadas. Através de técnicas de processamento de imagens é possível identificar, classificar e mensurar, por exemplo, a massa molecular de RNA, DNA, proteínas e enzimas, através da comparação com padrões pré-estabelecidos. Para auxiliar nessa tarefa existem alguns algoritmos já desenvolvidos bem como técnicas de Inteligência Artificial que realizam esse tipo de procedimento, com muita eficácia e precisão (LOPES, 1984).

Já o processamento digital de imagens consiste no melhoramento do aspecto visual de certas estruturas da imagem, além de proporcionar outros meios para que seja realizada a interpretação dos objetos presentes na imagem gerando novas saídas que podem ser utilizadas em processamentos posteriores (FILHO et al., 1999).

Este trabalho apresenta um algoritmo base para detecção do marcador e das colunas presentes nas imagens do gel de eletroforese e permitir que possam ser feitos novos estudos fundamentados nessas duas informações, visando criar uma ferramenta robusta e completa para auxílio do pesquisador.

1.2 Objetivos

1.2.1 Objetivo Geral

- Dar início a um sistema informatizado para o processamento automático da leitura, interpretação e análise de informações obtidas a partir de géis de eletroforese, fazendo uso de técnicas de processamento de imagens.

1.2.2 Objetivos específicos

- Iniciar o desenvolvimento de um sistema informatizado com base em técnicas de processamento de imagens com capacidade de detecção, interpretação e comparação de imagens de géis de eletroforese;
 - Possibilidade de carregamento de imagens digitais nos formatos BMP e JPG, coloridas;
 - Criar técnicas para localização das colunas nos géis de eletroforese e suas distancias correspondentes no gel.
 - Criar técnicas para a identificação do marcador de pares de bases.

1.3 Justificativa

A bioinformática vem na busca de soluções para enfrentar a explosão de dados e informações que são gerados constantemente nas áreas de biologia, biologia molecular, genética e afins.

No campo de pesquisa em análise de géis de eletroforese, existem poucas ferramentas que realizam esse tipo de análise, com isso a criação de uma ferramenta robusta se faz necessária.

Com relação a alguns problemas encontrados em programas que fazem a análise de imagens de géis de eletroforese, pode-se elencar:

- A produção de softwares que sejam de fácil manuseio e tenham uma interface intuitiva, voltada para os pesquisadores da área biológica;
- Problema na identificação de spots com posição sobreposta ou coincidente;
- Ruídos (pode-se citar "spots fantasmas").

A maior dificuldade envolvendo análise de eletroforese unidimensionais em géis de poliacrilamida e agarose, para os pesquisadores é a determinação da massa molecular dos fragmentos de DNA estudados.

Na UFPR existem alguns softwares que realizam a análise de géis, sendo um deles o LabWorks 4.0 (LabWorks Image Aquisition and Analysis Software, da UVP BioImaging Systems). Neste e em outros softwares os pesquisadores necessitam de certo conhecimento prévio para o manuseio dos mesmos, pois:

- São softwares muito complexos, no ponto de vista de sua usabilidade, já que necessitam que diversas configurações manuais sejam feitas pelo pesquisador.
- Em alguns casos o formato da saída dos dados fornecida pelo software precisa de várias adequações manuais, pois estes arquivos são somente compatíveis com o software que os gerou.

Devido a estes empecilhos, os pesquisadores sobrecarregam suas atividades com as anotações e procura de maneiras de reverter arquivos.

Criar uma ferramenta computacional que seja de fácil manuseio e que traga as informações necessárias sem muita dificuldade ou configurações manuais, traria

ao pesquisador uma otimização do seu trabalho, bem como a evolução e a economia de tempo em suas pesquisas.

1.4 Organização do trabalho

De forma resumida, esta dissertação será dividida em cinco capítulos, a saber:

No capítulo 2, é feita uma revisão bibliográfica de alguns conceitos de eletroforese e seus procedimentos, bem como o processamento de imagens, alguns conceitos e técnicas utilizados neste trabalho.

No capítulo 3, são mostrados os materiais e métodos usados nesse trabalho. Onde será abordado cada procedimento executado para chegarmos aos resultados alcançados.

No capítulo 4, são apresentados os resultados obtidos através da técnica desenvolvida.

Já no capítulo 5, são mostradas algumas comparações realizadas com outros softwares existentes no mercado que tem a mesma finalidade que o software desenvolvido e trazem algumas discussões sobre os géis utilizados, qualidades, aprimoramento na aquisição das imagens entre outras. Neste capítulo também é feita a conclusão do trabalho.

2 CAPITULO 2

2.1 *Revisão de literatura*

Nesta revisão da literatura são tratadas algumas definições de eletroforese, de processamento de imagens e uma explicação das técnicas utilizadas nesse trabalho.

2.1.1 Técnica de eletroforese de DNA em gel de poliacrilamida

Realizar análise de DNA envolvendo eletroforese é uma das técnicas mais presentes nos laboratórios que fazem diagnósticos e pesquisas. Como a molécula de DNA possui carga negativa em relação aos valores de pH neutro quando é aplicada em uma matriz de gel ligado a um campo elétrico, ocorre a migração dessas moléculas em direção ao pólo positivo (Figura 2.1). A rapidez da migração varia devido ao tamanho da molécula. Devido a isso em um determinado instante da eletroforese as moléculas que possuem tamanhos variados se encontram em posições diferentes na matriz (KONAREV, 1981).

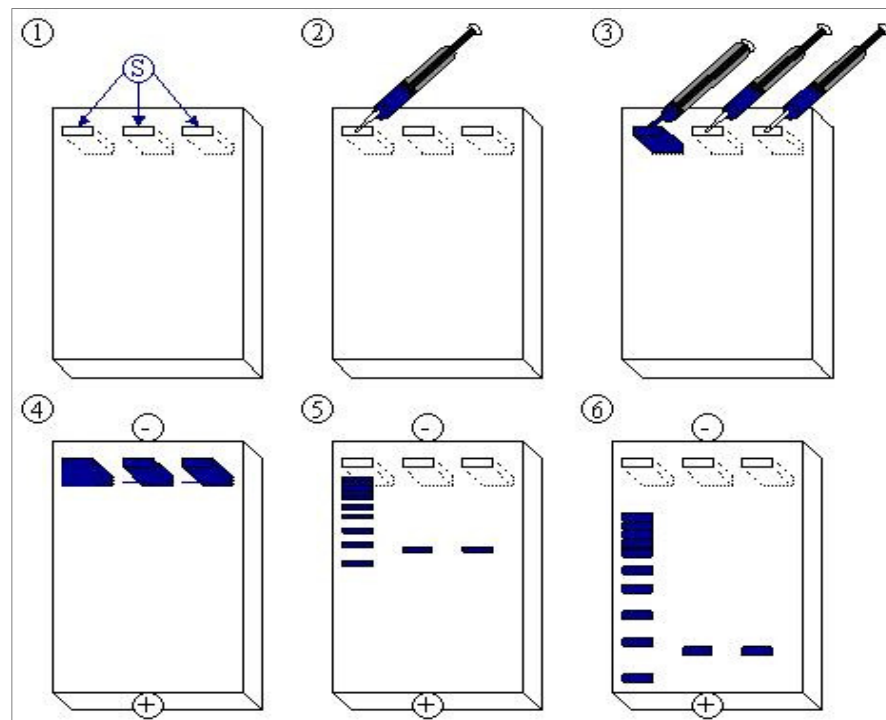


FIGURA 2.1 - TÉCNICA DE ELETROFORESE

1 - Gel; 2 - Aplicação Do Marcador De Pares De Bases; 3 - Aplicação Das Amostras; 4 - Indução De Corrente Elétrica; 5 - Deslocamento E Separação Dos Fragmentos; 6-Eletroforese Realizada.

FONTE: KONAREV (1981)

A matriz utilizada no procedimento para a aquisição das imagens deste trabalho é de poliacrilamida, e serve para fazer a desagregação de pequenos fragmentos, que podem ser de até 1kb (1 kb = 1000 pares de bases) (MARTINEZ, 2008; SHAW, 1969).

A coloração com o nitrato de prata, sal que se intercala as bases dos ácidos nucléicos é a mais comumente utilizada na visualização de DNA utilizando a técnica de eletroforese em géis de poliacrilamida (Figura 2.2). São detectadas quantidades iguais ou maiores que 10 ng de DNA e a proporção do tamanho da banda obtida pela revelação do gel, está relacionada com a concentração de DNA na amostra (SHAW, 1969).

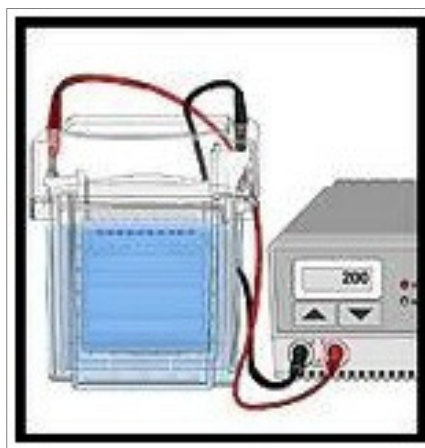


FIGURA 2.2 - APARELHOS UTILIZADOS NA TÉCNICA DE ELETROFORESE VERTICAL EM GEL DE POLIACRILAMIDA.

FONTE: <http://biomedicinabrasil.blogspot.com/2010/10/eletroforese.html>

Conhecidas como bandas de DNA (Figura 2.3), elas são originadas pelas moléculas de DNA inseridas em um dos poços feitos no gel e migram na matriz de poliacrilamida assumindo a forma do poço (BLOCK et al., 1958).

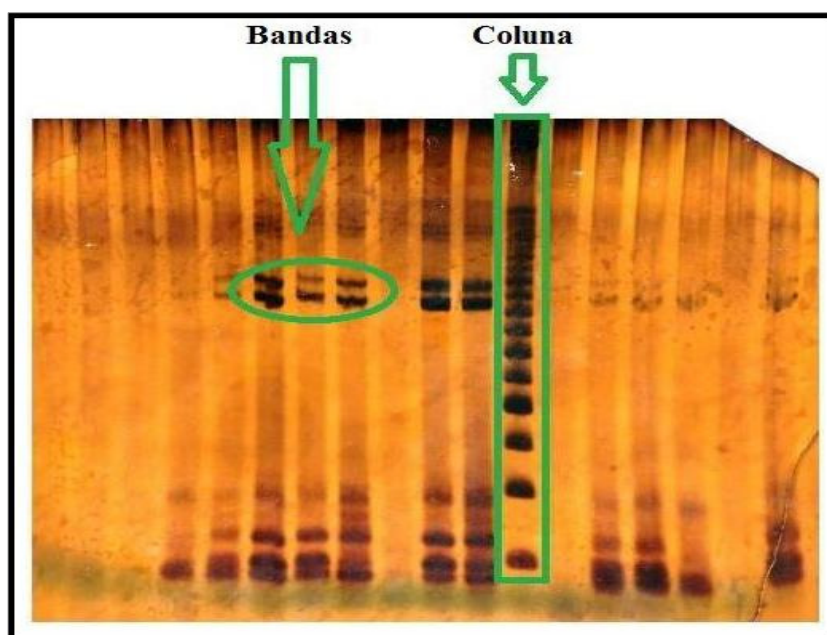


FIGURA 2.3 – BANDAS E COLUNAS

Indica as bandas que ficaram mais evidentes com a coloração com prata e na posição vertical estão as colunas, onde as moléculas de DNA migram conforme seu peso molecular.

FONTE: O autor (2011)

Após a observação dos resultados, normalmente é utilizado um scanner comum para realizar a digitalização e posterior análise da imagem do gel de eletroforese. Para esse trabalho foi usado essa técnica, sendo digitalizadas imagens com 300 d.p.i., como mostra a Figura 2.4.

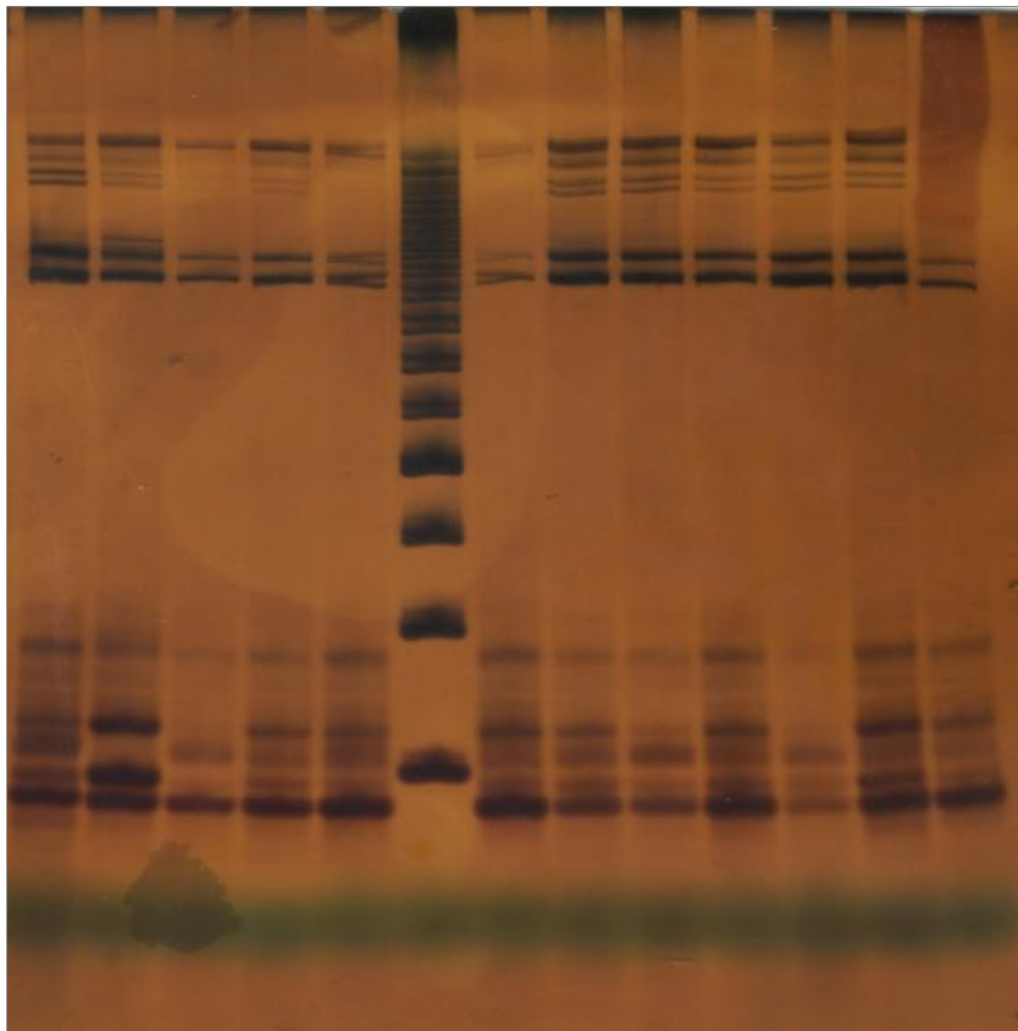


FIGURA 2.4 - EXEMPLO DE IMAGEM UTILIZADA NESSE TRABALHO

FONTE: O autor (2011).

Existem vários sistemas para análise utilizando a eletroforese em gel de poliacrilamida, tais como:

- Eletroforese em condições não-desnaturantes: normalmente empregada quando a análise da atividade biológica, como por exemplo, a atividade enzimática de uma proteína, deve ser testada. Nesse caso, a manutenção da

proteína na sua conformação nativa se faz necessária, evitando-se o uso de agentes desnaturantes;

- Eletroforese isoeletrica: empregada para a determinação da carga elétrica líquida das proteínas. A uréia é usada como agente dissociante em combinação com um gradiente de pH que é gerado por uma mistura de anfólitos, oligômeros de baixa massa molecular (300-600 Da) que determinam uma ampla faixa de ponto isoeletrico (PI) ao longo do gel. Nesse sistema de eletroforese, as proteínas migram até atingir o seu PI. Nesse ponto, as moléculas de proteínas da amostra não mais se deslocam e permanecem imobilizadas como bandas no gel;
- Eletroforese bidimensional: o sistema de maior poder de resolução, uma vez que em primeira etapa a separação das moléculas é baseada na carga elétrica das proteínas (focalização isoeletrica) e, na segunda dimensão, pela massa molecular das mesmas (em condições desnaturantes). Tem-se então uma separação das moléculas de proteína em função de dois parâmetros físico-químicos importantes (carga e tamanho), conferindo maior poder de resolução na análise das proteínas presentes na amostra (AZEVEDO, 2003).

A técnica utilizada nesse trabalho é resultante da eletroforese unidimensional em condições desnaturantes (SDS-PAGE: Sodium Dodecyl Sulphate – Polyacrilamide Gel Eletrophoresis).

O sistema SDS-PAGE é o mais comumente usado para determinação da massa molecular de proteínas ou DNA, apresentando uma precisão de até 95 %. A condição desnaturante provida pelo SDS (detergente aniônico forte) usado em combinação com um agente redutor (geralmente o β -mercaptoetanol) e calor (aquecimento da amostra a 100°C) (AZEVEDO, 2003). Nessas condições, as proteínas da amostra tornam-se completamente dissociadas/desagregadas, preservando apenas sua estrutura primária.

Os polipeptídios desnaturantes ligam-se ao SDS e tornam-se carregados negativamente. A migração do complexo SDS-proteína ou DNA através do gel de poliacrilamida dá-se em função do tamanho do polipeptídio, uma vez que a quantidade de SDS ligado a proteína independe de sua seqüência primária de aminoácidos e é proporcional a sua massa molecular (tamanho). Na saturação, cerca de 1,4g de SDS é ligado por grama de proteína, o que corresponde a cerca de

uma molécula de SDS para cada dois resíduos de aminoácidos. Nessas condições, a mobilidade relativa de uma proteína varia linearmente com o logaritmo de sua massa molecular (AZEVEDO, 2003).

Na maioria dos casos, o sistema SDS-PAGE é realizada com um sistema de tampão descontínuo, no qual o tampão nos reservatórios da cuba de eletroforese apresenta pH e força iônica diferente do tampão usado para confeccionar o gel. Quando a amostra é aplicada, os complexos de SDS-proteína são arrastados por uma fronteira de íons em movimento, que é criada quando a corrente elétrica passa através dos eletrodos, do pólo negativo em direção ao pólo positivo. A combinação do sistema descontínuo de tampão, com a grande porosidade do gel concentrador promove o efeito de empacotamento da amostra, o qual é crucial para a alta resolução desse sistema de separação (AZEVEDO, 2003).

Após a separação pelo sistema SDS-PAGE, as proteínas ou DNA podem ser visualizadas segundo diversas metodologias de coloração, detecção imunológica ou radioativa. Normalmente são usados na metodologia de coloração o corante Coomassie Blue e sais de prata (soluções amoniacaís de prata ou nitrato de prata). Os diversos processos de coloração pela prata baseiam-se na redução diferencial dos íons prata ligados as cadeias laterais dos aminoácidos. A coloração pela prata é cerca de cem mil vezes mais sensível que a coloração pelo Coomassie, podendo detectar moléculas em uma quantidade de 0,1 a 1,0 ng de proteína em uma única banda (AZEVEDO, 2003).

2.1.2 Marcadores moleculares

Os marcadores moleculares estão cada vez mais sendo utilizado, devido apresentar vantagens sobre os marcadores morfológicos, que eram utilizados até a década de 60, nos estudos de genética e melhoramento (FERREIRA e GRATTAPAGLIA, 1998).

A disponibilidade de marcadores morfológicos era basicamente exclusiva a poucas espécies de plantas utilizadas como sistemas modelo para o estudo da genética. A mudança profunda começou com o desenvolvimento de marcadores

isoenzimáticos e, posteriormente, melhorados com o surgimento de técnicas que detectam polimorfismo genético diretamente no DNA. Atualmente, um número ilimitado de marcadores moleculares altamente polimórficos pode ser conseguido em qualquer organismo fazendo uso de diversas técnicas (FERREIRA e GRATTAPAGLIA, 1998).

Os marcadores moleculares mais comumente utilizados podem ser classificados em dois grupos, de acordo com a metodologia usada para identificá-los: hibridização ou amplificação de DNA. No caso dos identificados por hibridização estão os marcadores RFLP (Restriction Fragment Length Polymorphism) e minissatélites ou locos VNTR (Variable Number of Tandem Repeats). E para os evidenciados por amplificação são os marcadores do tipo RAPD (Random Amplified Polymorphic DNA), microssatélite e AFLP (Amplified Fragment Length Polymorphism) (FERREIRA e GRATTAPAGLIA, 1998).

2.1.3 Microssatélites

Espaços pequenos de seqüências simples sucedem como elementos altamente repetitivos em tandem em todos os genomas eucariotos e em partes também em procariotos e eubactérias (TAUTZ, 1989). Estas seqüências, são conhecidas como microssatélites ou SSR ("Simple Sequence Repeats"), possuem repetições das unidades nos diferentes locos que diversificam de um a cinco nucleotídeos em tamanho. Pelo motivo da grande variabilidade no número de repetições da maioria dos locos, os microssatélites são amplamente utilizados em aplicações forenses, estudos genéticos de populações, mapeamento genético e físico de genomas e, com isso, são o suporte principal de análises genômicas em diferentes organismos (RUBINSZTEIN et al., 1995).

Estes locos altamente polimórficos, são amplificados via PCR (Reação em Cadeia da Polimerase), também foram denominados "STMS – Sítios de Microssatélites Marcados por Seqüências" e formam a classe mais polimórfica de marcadores moleculares que existem hoje em dia, podendo ser rapidamente isolados em bibliotecas genômicas. Cada segmento amplificado de tamanho

diferente representa um alelo diferente do mesmo loco (FERREIRA e GRATAPAGLIA, 1998).

Os microssatélites são geralmente mono-, di-, tri- ou tetranucleotídeos em um loco simples. Muitos locos de microssatélite estão localizados entre genes ou dentro de íntrons, sendo extremamente abundantes (ENGEL et al., 1996). Entretanto apesar de muito raros nas regiões codificadoras, no genoma humano são encontrados alguns tipos de repetições trinucleotídicas nestas regiões, sendo a causa de algumas doenças (HANCOCK, 1999; RUBINSZTEIN et al., 1995).

Para explicar as altas taxas de mutações nos microssatélites existem dois modelos: O crossing-over que não são iguais entre as moléculas de DNA, resultando da recombinação entre cromossomos homólogos que não foram alinhados de maneira correta e o mecanismo de slippage que é realizado durante a replicação do DNA, que é o mais aceito atualmente, ocorrendo em regiões do DNA que contêm seqüências curtas repetidas em tandem, proporcionado por uma elevada taxa de erros de pareamento decorrentes do escorregamento da fita durante a replicação do DNA (EISEN, 1999).

Devido à distribuição preferencial nas regiões não codificadoras, os microssatélites podem não sofrer ação da seleção natural, o que os torna um marcador seletivamente neutro muito útil para estudos da genética de populações naturais (GONELA, 2003).

No entanto, alguns estudos demonstram que eles podem apresentar papel funcional com elementos codificantes ou reguladores, mostrando que a utilidade dos microssatélites vai além do papel de marcadores para estudos de genética de populações, relações evolutivas e mapeamento gênico. Deste modo, um loco poderia mostrar um desvio da neutralidade se o microssatélite estivesse ligado a uma região genômica a qual é alvo da seleção natural (MACHADO, 2003; GONELA, 2003).

Devido a sua expressão codominante e ao multialelismo, os microssatélites são os marcadores que apresentam o mais elevado conteúdo de informação de polimorfismo (PIC) e, por isso, qualquer população segregante pode ser utilizada como população referência para estudos de ligação e mapeamento, não sendo mais necessário fazer a escolha da população com base na maximização da distância

genética, e sim visando a população mais informativa do ponto de vista das características biológicas ou econômicas de interesse (FERREIRA e GRATTAPAGLIA, 1998).

A descoberta de que locos de microssatélites dinucleotídeos são conservados através dos genomas de mamíferos (SUN e KIRKPATRICK, 1996) indicou que iniciadores desenvolvidos para amplificar repetições em uma espécie podem amplificar locos em outras espécies. A utilização de iniciadores heterólogos (amplificam dentro de um mesmo gênero, família ou ordem) pode eliminar tempo e despesas envolvidos no isolamento de marcadores microssatélites para cada espécie de interesse tais como, construção de biblioteca genômica de DNA, seqüenciamento do DNA e determinação dos iniciadores (ENGEL et al., 1996).

2.2 Processamento de imagens

Ao processar uma imagem, esta é simbolizada por $f(x, y)$, que é uma função bidimensional que tem nas coordenadas (x, y) seu nível de luz, como pode ser observado na Figura 2.5. Este nível se encontra em um intervalo que vai do zero ao infinito, devido a luz ser considerada uma forma de energia (MORAES, 1990).

Entretanto, o que o olho humano consegue enxergar, está relacionado a dois fatores elementares, um deles é o volume de luz que recai em um objeto e a outra é o volume dessa luz refletida por esse objeto (FILHO, 1999). O computador armazena a imagem de forma numérica, por isso a imagem é tratada como uma matriz de $N \times M$, do qual os índices de linhas e de colunas reconhecem um ponto na imagem, e o correspondente valor do elemento da matriz reconhece o nível de cinza presente nesse ponto (CONCI et al., 2007).

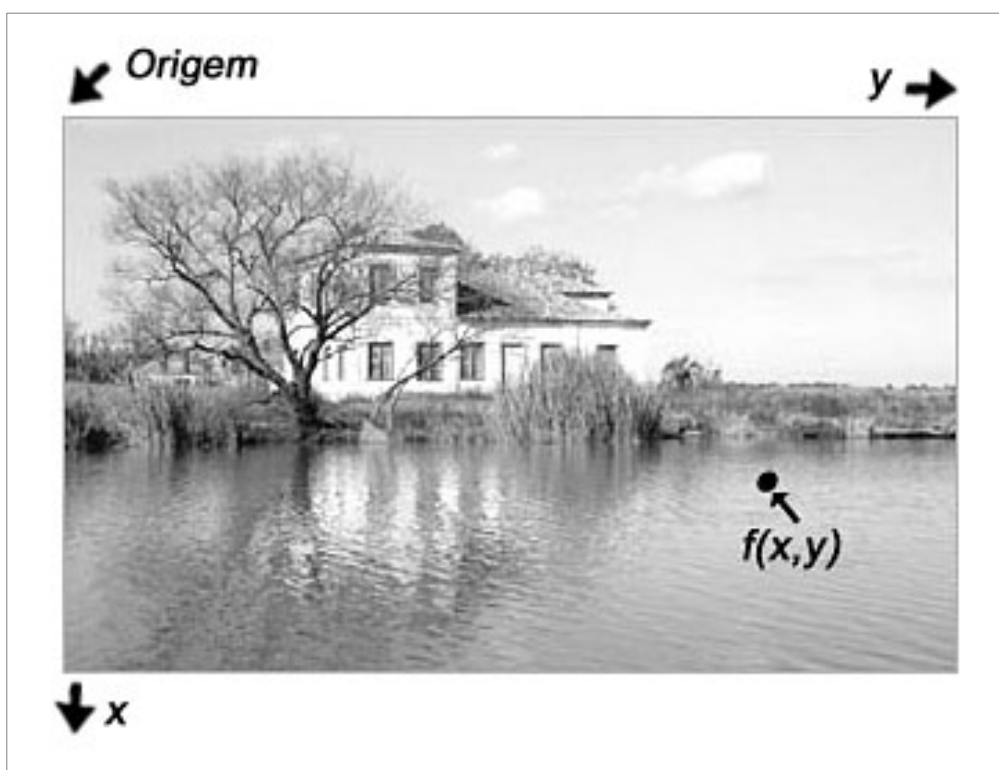


FIGURA 2.5 - EXEMPLO DA CONVENÇÃO DOS EIXOS PARA REPRESENTAÇÃO DE IMAGENS DIGITAIS

FONTE: MORAES (1990).

A matriz possui informações das coordenadas espaciais da imagem que foi digitalizada, conhecida como amostragem, e também as características da sua quantificação (GONZALES, 2000).

A amostragem se remete a resolução da imagem e suas coordenadas espaciais, podendo ser, por exemplo, de 256x256 pixels, 512x512 pixels, entre outras possíveis, portanto quanto maior for a quantidade de pixels da imagem, melhor e sua resolução (MATARREDONA, 1994).

A quantificação se refere a profundidade dos níveis de cinza, que usualmente são de 16, 32, 64, 128 ou 256 níveis de cinza (ROSENFELD, 1993).

Na Figura 2.6 mostra a relação de Resolução x Profundidade.

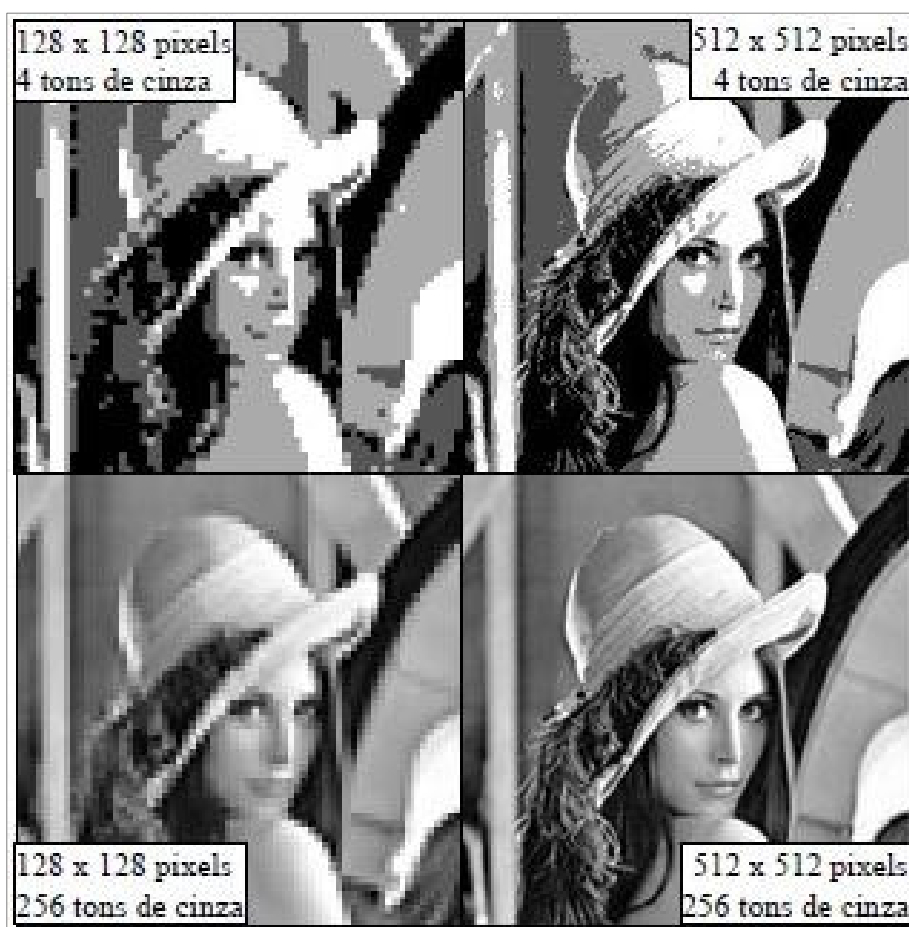


FIGURA 2.6 - RESOLUÇÃO X PROFUNDIDADE

FONTE: ROSENFELD (1993).

Um dos modelos mais utilizados em processamento de imagens é o monocromático, neste modelo cada ponto da imagem é representado de apenas uma banda espectral (as imagens coloridas possuem três ou mais bandas espectrais). Para representar uma imagem monocromática, necessita-se armazenar apenas 1 bit por pixel, ou seja, o pixel é preto (bit=0) ou branco (bit=1) (CONCI et al., 2007).

Pode-se citar um exemplo de aplicação, que é fácil de ser explicada sem muito conhecimento preliminar de conceitos de imagens, é a utilização de técnicas de processamento de imagens para separar e classificar objetos em uma imagem (CASTLEMAN, 1995).

A Figura 2.7 demonstra que o objetivo principal é gerar um resultado a partir do domínio do problema, através de processamento de imagens. Nesse caso, o

domínio do problema está em analisar a imagem, identificar os objetos, e separar os objetos do fundo da imagem. Dessa forma, a saída esperada é a classificação ou separação dos objetos em seus respectivos grupos.

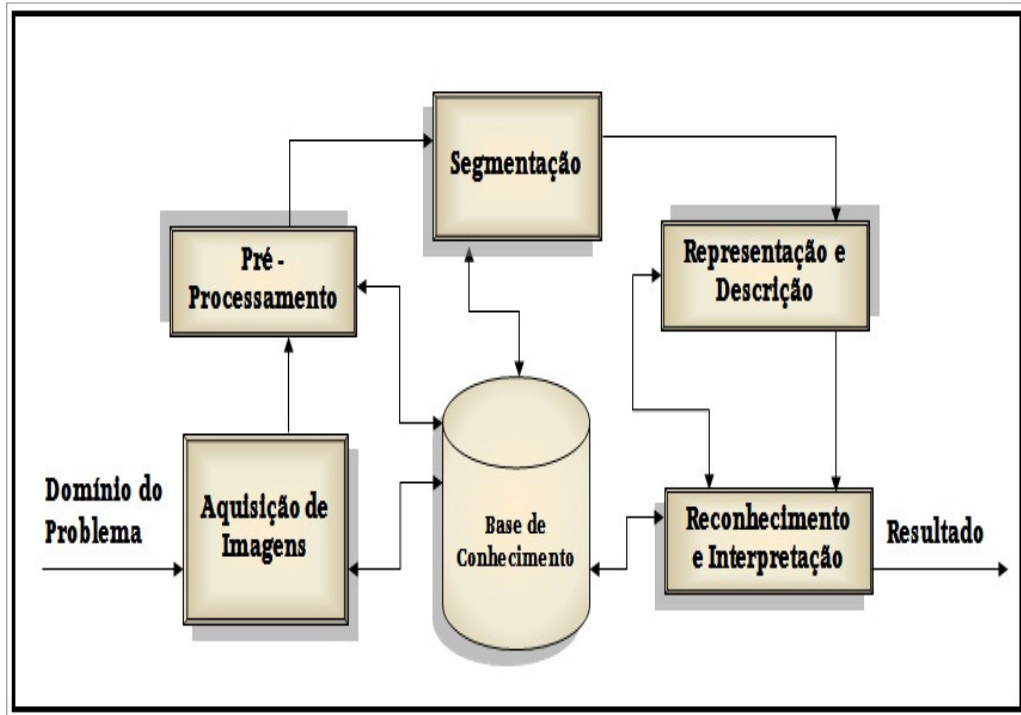


FIGURA 2.7 - ETAPAS NO PROCESSAMENTO DE IMAGENS

FONTE: O autor (2011).

A aquisição da imagem é o primeiro passo no processo, ou seja, obter uma imagem digital. Nesse processo é necessário um dispositivo de aquisição de imagem que seja sensível ao sinal proveniente de alguma fonte de radiação, por exemplo, calor, raios X, luz visível, entre outros, além da transformação deste sinal em um sinal digital (JAIN, 1989).

Depois de ser obtida a imagem digital, a etapa a seguir é pré-processar essa imagem. A premissa utilizada no pré-processamento visa melhorar a imagem para que seja possível ampliar as chances para o êxito nas próximas etapas. Nesse caso, o pré-processamento faz uso de técnicas que realizam o realce de contraste, eliminação de ruídos, etc. (CASTLEMAN, 1995).

A próxima etapa está relacionada com a segmentação da imagem, que separa uma imagem de entrada em partes ou objetos. Geralmente, a técnica de segmentação automática está entre as tarefas mais difíceis empregadas no

processamento de imagens digitais. Por sua vez, quando a técnica utilizada para realizar a segmentação é bem elaborada, ela favorece potencialmente a resolução do problema de imagens ruins. Entretanto, a utilização de algoritmos de segmentação de imagens mal desenvolvidos ou fracos faz com que ocorram prejuízos no processamento (JAIN, 1989).

O resultado da etapa de segmentação é formado normalmente por dados em sua maioria de pixels, que representam todos os pontos existentes na imagem, como o limite de uma região. Nos dois casos, é essencial fazer a conversão dos dados para uma forma correta para ser utilizada no processamento computacional.

Na fase de seleção de características ou representação, são extraídas algumas características que derivam em uma informação quantitativa útil para a distinção entre os tipos de objetos (CASTLEMAN, 1995).

Por fim, a etapa que implica no reconhecimento, rotula os objetos, fundamentado na informação provida pela fase de representação. A acepção está relacionada com a atribuição de significado a um conjunto de objetos já rotulados (CASTLEMAN, 1995).

Exemplificando de maneira simples, a identificação de um objeto, que pode ser uma "bola", necessita da associação dos descritores para esse objeto com o rótulo "bola". A acepção visa conceder significado a um conjunto de objetos já rotulado (JAIN, 1989).

2.2.1 Histograma

O histograma de qualquer imagem é uma representação gráfica da distribuição dos pixels entre os seus níveis de cinza. Tem a finalidade de proporcionar uma visão global da distribuição, mas nada diz sobre seu conteúdo. Tem como eixo horizontal a escala de tons de cinza e no eixo vertical representa as ocorrências dos tons de cinza na imagem (CARVALHO, 2003)

RABBANI (1992) define o histograma da seguinte maneira: Em uma imagem $f(x, y)$ de dimensão $N \times M$, o histograma é dado pela função:

$$p(r) = \frac{n_r}{n}$$

FIGURA 2.8 - FUNÇÃO DO HISTOGRAMA DE UMA IMAGEM $F(X, Y)$ DE DIMENSÃO $N \times M$.

FONTE: RABBANI (1992)

Onde r representa um valor inteiro qualquer no intervalo de 0 a 255, o qual representa os níveis de cinza de uma imagem, n_r representa o numero de vezes que este valor r aparece na imagem e n é o total de pixels da imagem $f(x, y)$.

O histograma é uma ferramenta muito útil na fase de pré-processamento, pois fornece uma visão estatística sobre a distribuição dos pixels, sobre o contraste da imagem e os níveis de iluminação (ANDREWS et al., 1972). Ele também é muito difundido na fase de segmentação, principalmente em técnicas que se utilizam da similaridade entre os pixels.

O histograma é utilizado com freqüência como sendo uma distribuição estatística dos pixels ("luminância") na imagem, como, por exemplo, no caso das técnicas que o utilizam para calcular a entropia da imagem (COLTUC et al., 2006).

A Figura 2.9, mostra duas imagens e seus respectivos histogramas:

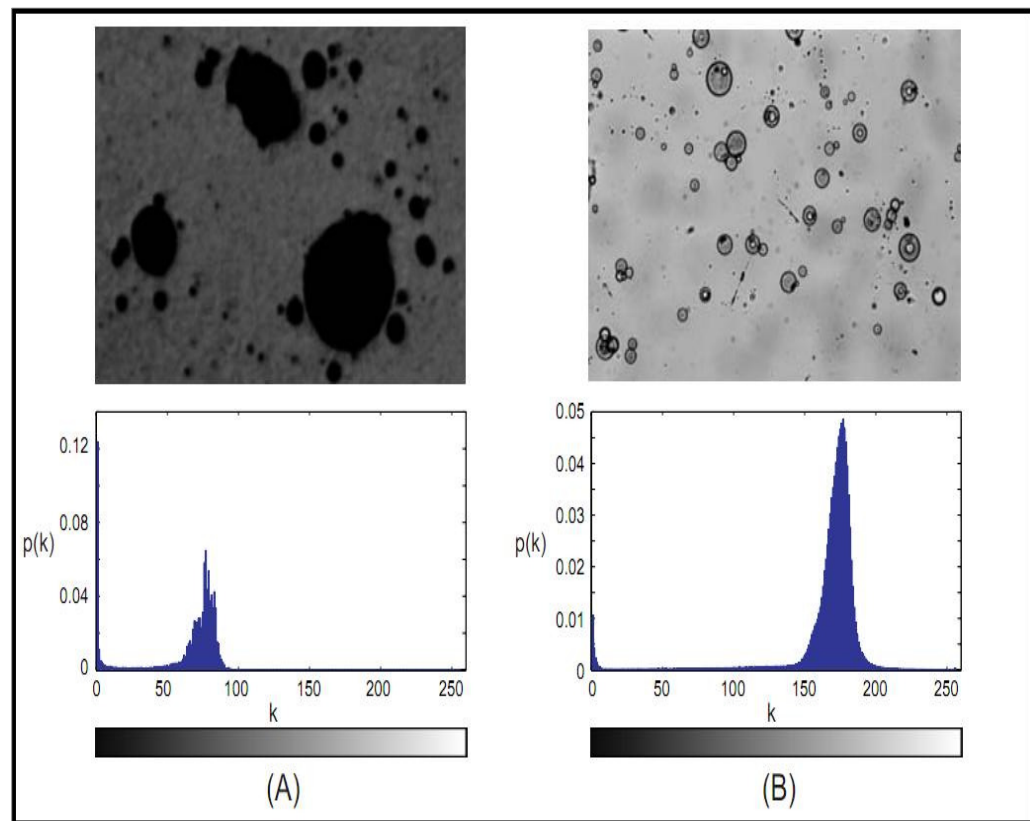


FIGURA 2.9 - EXEMPLO DE HISTOGRAMAS

A - É Uma Imagem De Baixo Contraste, B - Possui Um Maior Contraste.

FONTE: COLTUC et al. (2006).

2.3 Filtros de Suavização

Em grande parte das aplicações, é notável a necessidade de suavizar uma imagem na medida em que suas arestas são preservadas. Métodos para realizar a suavização de imagens como, por exemplo, o filtro passa-baixa, que tem como princípio não considerar a variação de intensidade da imagem, visam borrar as arestas.

2.3.1 Filtro Gaussiano

Esse filtro é comumente empregado para suavização de imagens, mas tem como característica principal a não preservação das arestas, por não considerar a diferença das intensidades.

O filtro Gaussiano possui dois parâmetros, a área da janela é um valor usado para o desvio padrão máximo sigma. Seu procedimento nos faz lembrar o filtro passa-baixa, que nada mais é do que um suavizador de imagens (DURAND, 2002).

A intensidade da suavização na imagem está ligada ao desvio padrão sigma, ou seja, quanto maior for o valor de sigma, maior será a suavização da imagem, não estando muito relacionado com a característica referente a área da janela. Quanto maior for o valor de sigma, maior será o número de pixels do qual o valor é distinto de zero, fazendo com que os pixels vizinhos tenham maior importância em cada ponto, assim efetuando uma suavização extra na imagem (KOPF, 2007).

A definição do filtro Gaussiano (GB) é dada por:

$$G_{\sigma}(x) = \frac{1}{2\pi \sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

FIGURA 2.10 – DEFINIÇÃO DO FILTRO GAUSSIANO (GB).

FONTE: KOPF (2007).

Onde σ é o desvio padrão. É assumido que a distribuição tem média zero, ou seja, está centrada em $x=0$. A distribuição é ilustrada na Figura 2.11.

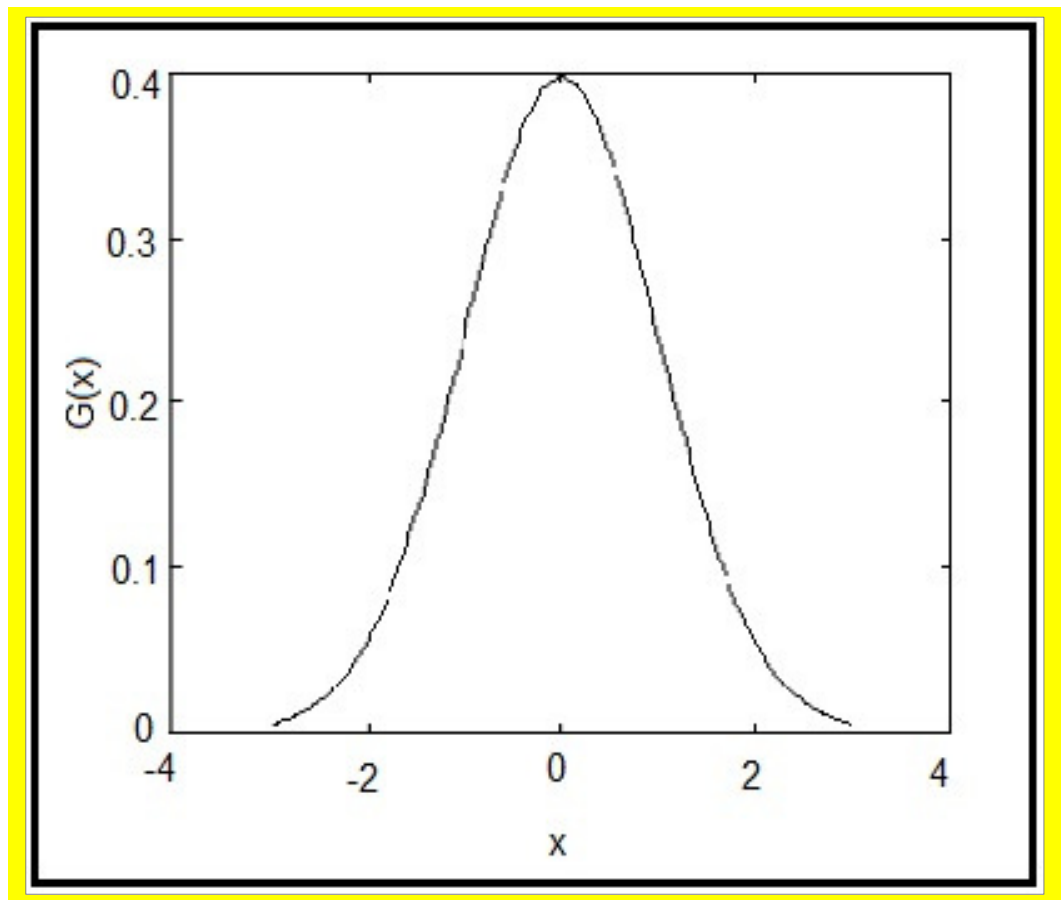


FIGURA 2.11 – DISTRIBUIÇÃO GAUSSIANA COM MÉDIA ZERO E DESVIO UM, SIGMA = 1, NA EQUAÇÃO DA FUNÇÃO GAUSSINA

FONTE: KOPF (2007).

A Figura 2.12, apresenta o resultado com o uso do filtro gaussiano com diferentes valores para o sigma σ .

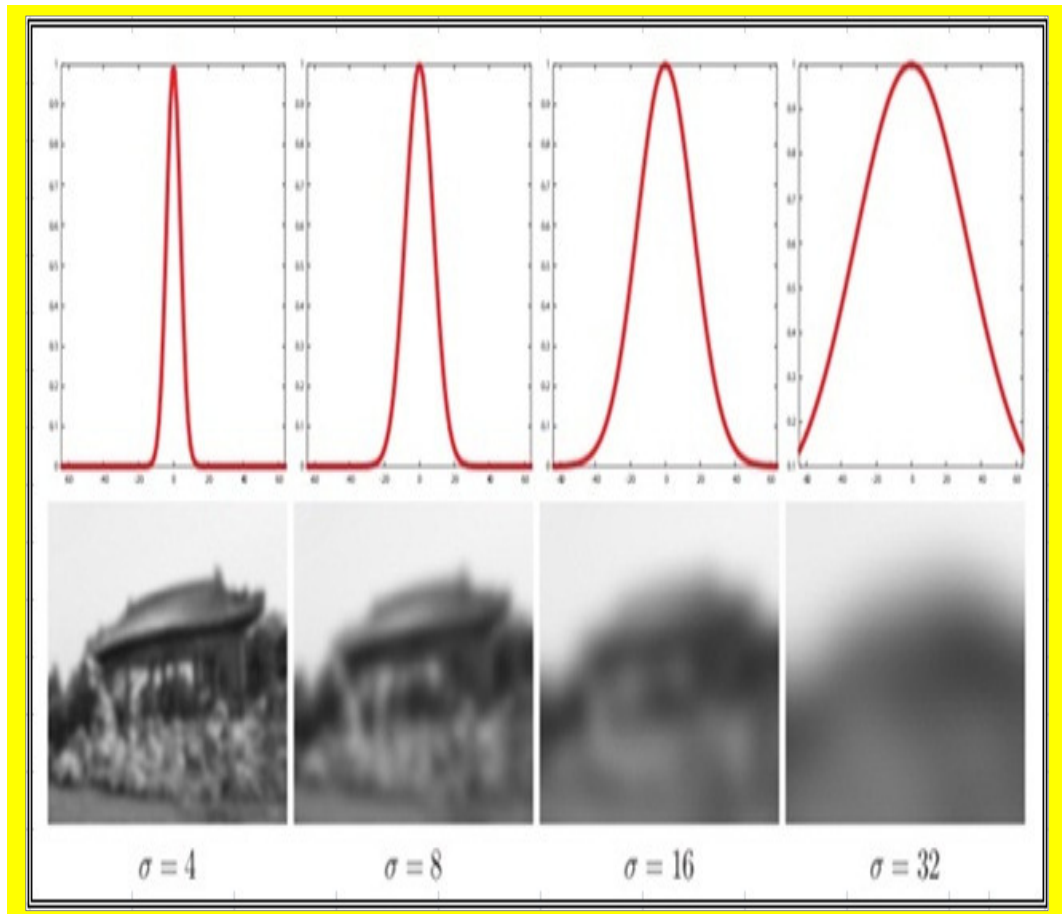


FIGURA 2.12 - RESULTADO COM O USO DO FILTRO GAUSSIANO COM DIFERENTES VALORES PARA O SIGMA.

FONTE: KOPF (2007).

2.3.2 Filtro Bilateral

O filtro bilateral foi incorporado como um filtro não-linear que visa combinar filtros no domínio e no espaço de cores da imagem. Exemplificando: para uma determinada imagem de entrada T , consegue-se uma imagem de saída T' (ELAD, 2002).

É uma técnica utilizada para suavização de imagens visando a preservação de arestas. O seu uso teve uma rápida aceitação na área de processamento de imagens, se tornando corriqueiro.

Algumas aplicações que fazem o uso dessa técnica são para edição de textura, edição de iluminação, remoção de ruído, etc. (PETSCHNINGG et al., 2004).

As principais vantagens de se utilizar o filtro bilateral são as seguintes:

- A fórmula utilizada para sua implementação é simples, sendo que cada pixel é modificado pela média de seus vizinhos. Com isso fica mais fácil entender o seu comportamento, podendo assim ajustar e elaborar seu funcionamento para uma aplicação singular (PARIS, 2007).
- Faz uso de apenas dois parâmetros (G_{σ_s} e G_{σ_r}): um indica o espaço e o outro o range das particularidades a serem conservadas.
- Com parâmetros de fácil configuração, ele pode ser utilizado de forma automatizada.
- Pode ser calculado de maneira ágil para imagens de grande porte e se utilizado um hardware gráfico, é útil para imagens em tempo real (PETSCHNINGG et al., 2004).

A principal diferença do filtro bilateral em relação ao filtro gaussiano é que o primeiro utiliza a variação das intensidades para poder fazer a preservação das arestas. No caso do filtro bilateral, se dois pixels estão próximos, eles não só estão ocupando posições espacialmente próximas, mas também são análogos na escala fotométrica (PARIS, 2007; PETSCHNINGG et al., 2004).

A definição do filtro bilateral (BF) é dada por:

$$BF[I]_{\mathbf{p}} = \frac{1}{W_{\mathbf{p}}} \sum_{\mathbf{q} \in \mathcal{S}} G_{\sigma_s}(\|\mathbf{p} - \mathbf{q}\|) G_{\sigma_r}(I_{\mathbf{p}} - I_{\mathbf{q}}) I_{\mathbf{q}}$$

FIGURA 2.13 – DEFINIÇÃO DO FILTRO BILATERAL (BF).

FONTE: PETSCHNINGG et al. (2004).

Sendo que W_p é um elemento de normalização:

$$W_p = \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(I_p - I_q)$$

FIGURA 2.14 – WP COMO SENDO UM ELEMENTO DE NORMALIZAÇÃO

FONTE: PETSCHNINGG et al. (2004).

A equação acima é considerada uma média normalizada dos pesos, sendo que G_{σ_s} é uma gaussiana espacial que tem seu valor decrescente conforme a distância do pixel, e G_{σ_r} é uma gaussiana no espaço de cores.

2.3.2.1 Parâmetros

O filtro Bilateral é ajustado fazendo uso de dois parâmetros: σ_s e σ_r . A imagem mostrada na Figura 2.15, exhibe o efeito dos parâmetros espaciais (σ_s) e do espaço de cor (σ_r), e uma confrontação com o filtro gaussiano. Lembrando que quanto mais adjacente de zero é o peso, menor será a suavização (KOPF, 2007; DURAND, 2002).

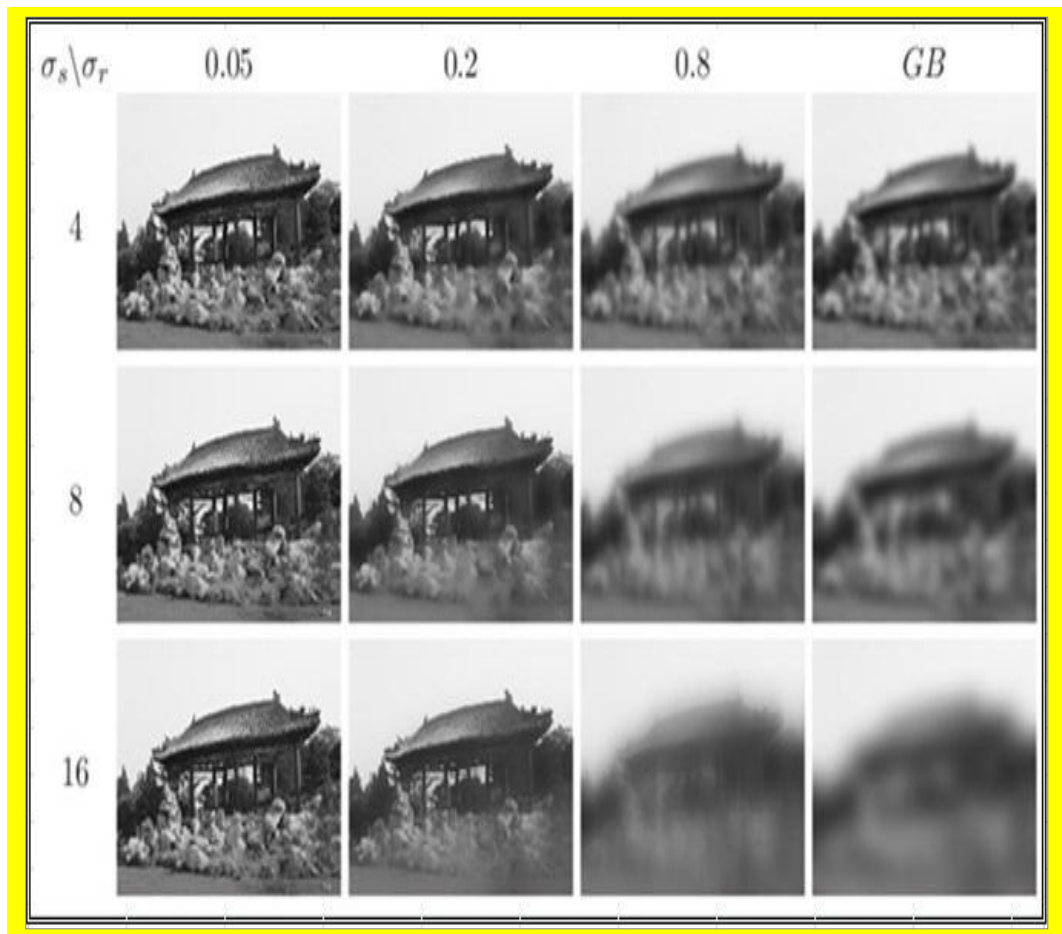


FIGURA 2.15 – EFEITO DOS PARÂMETROS ESPACIAS (SIGMA S) E DO ESPAÇO DE COR (SIGMA R) E UMA CONFRONTAÇÃO COM O FILTRO GAUSSIANO

FONTE: DURAND (2002).

Quando aumentamos o parâmetro do espaço de cor " σ_r ", o filtro bilateral tende a ficar mais próximo do filtro gaussiano, devido seu espaço de cor ser plano, ou seja, próximo de ser uma constante sobre a distância de intensidade da imagem (ELAD, 2002).

Quando aumentamos o parâmetro espacial " σ_s ", detalhes maiores presentes na imagem são suavizados. Vale ressaltar que o filtro bilateral tem uma importante particularidade, onde os pesos são multiplicados, o que resulta em quanto mais adjacente de zero for o peso, a suavização se torna nula, ou seja, nenhuma suavização acontece (DURAND, 2002).

2.4 Segmentação

A segmentação ou limiarização, de forma simplificada, consiste em separar a imagem como um todo nas partes que a constituem e que se diferenciam entre si. É usual denominar "objetos" da imagem os grupos de pixels de interesse, ou que fornecem alguma informação para o processamento de imagem. Da mesma forma, a denominação "fundo" da imagem é utilizada para o grupo de pixels que podem ser desprezados ou que não tem utilidade no processamento de imagem. Essas denominações "objeto" e "fundo" possuem uma conotação bastante subjetiva, podendo se referir a grupos de pixels que formam determinadas regiões na imagem sem que representem um objeto, de modo literal, presente na imagem processada (CONCI et al., 2007).

A segmentação, em relação a todas as etapas do processamento de imagens é considerada a etapa mais crucial do tratamento da informação. E nesta etapa que são definidas as regiões de interesse para processamento e análise posteriores.

Devido a este fato, quaisquer erros ou distorções presentes nesta etapa irão refletir nas próximas etapas, com isso irá produzir ao término do processo resultados não desejados que possam contribuir de forma negativa para a eficiência de todo o processamento (ANDREWS et al., 1972).

Vale lembrar que não existe um modelo formal para a segmentação de imagens. A segmentação é um processo empírico e adaptativo, que visa se adequar as características particulares de cada tipo de imagem e aos objetivos que se pretende alcançar. Mesmo existindo uma grande diversidade de técnicas de segmentação de imagens, ainda existe atualmente, um interesse maior no estudo e desenvolvimento de novas técnicas (COLTUC et al., 2006; CARVALHO, 2003)

As técnicas de segmentação utilizam duas abordagens principais: a similaridade entre os pixels e a descontinuidade entre eles. A técnica que se baseia em similaridade mais utilizada é a chamada binarização.

A binarização de imagens ou *Thresholding* é uma técnica eficiente e simples do ponto de vista computacional, sendo, portanto largamente utilizada em sistemas

de visão computacional. Esta segmentação é utilizada quando as amplitudes dos níveis de cinza são suficientes para caracterizar os "objetos" presentes na imagem.

Na binarização, um nível de cinza é considerado como um limiar de separação entre os pixels que compõem os objetos e o fundo. Nesta técnica, se obtém como saída do sistema uma imagem binária, ou seja, uma imagem com apenas dois níveis de luminância: preto e branco. A determinação deste limiar de modo otimizado para segmentação da imagem é o objetivo principal dos diversos métodos de binarização existentes (GONZALEZ, 2000; CASTLEMAN, 1995).

As técnicas que fazem uso da descontinuidade entre os pixels procuram determinar variações abruptas do nível de luminância entre pixels vizinhos. Estas variações, normalmente, permitem detectar o grupo de pixels que delimitam os contornos ou bordas dos objetos na imagem. A técnica de segmentação baseada em descontinuidade mais utilizada é a chamada detecção de bordas (CASTLEMAN, 1995; RABBANI, 1992).

Na Figura 2.16, se traz dois exemplos de segmentação, uma binarização e uma detecção de bordas.

A etapa de pós-processamento geralmente é a etapa que sucede a segmentação. É nesta etapa que os principais defeitos ou imperfeições da segmentação são devidamente corrigidos. Normalmente, estes defeitos da segmentação são corrigidos através de técnicas de Morfologia Matemática, com a aplicação em seqüência de filtros morfológicos que realizam uma análise quantitativa dos pixels da imagem, entre outras técnicas (JAIN, 1989).

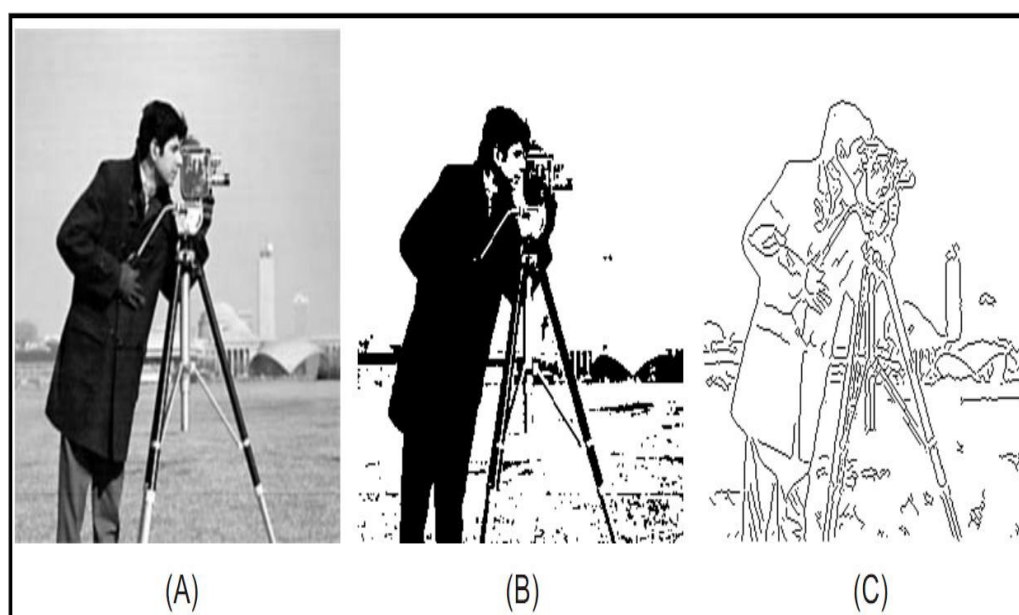


FIGURA 2.16 - EXEMPLO DE SEGMENTAÇÃO

A - É a imagem original em níveis de cinza, B – Imagem segmentada por binarização e C - Imagem segmentada por detecção de bordas.

FONTE: CASTLEMAN (1995).

2.5 Morfologia Matemática Binária

A morfologia matemática está relacionada com o estudo e análise de imagens usando operadores não lineares. Foram descobertos por Georges Matheron e Jean Serra que focavam seus estudos em estruturas dentro de imagens binárias. (BANON et al., 2007).

Com a utilização da morfologia, são executadas algumas atividades no processamento de imagens que visam o aumento de qualidade, como: a restauração de imagens, detecção de falhas, análise da textura, análise particular, análise de formas, compreensão, dentre outras. Essas ferramentas são utilizadas em várias áreas como robótica, inspeções, microscopia, biologia, bioinformática, entre outras (DOUGHERTY et al., 2003).

O princípio da morfologia matemática se fundamenta na geometria e na topologia de um conjunto não conhecido de uma imagem.

Basicamente a idéia é percorrer uma imagem com um elemento estruturante e quantificar a maneira com que este se encaixa ou não naquela parte da imagem. Caso seja afirmativo, o local é demarcado, isto indica onde o elemento estruturante se encaixou na imagem. Com isso, é possível extrair informações importantes, como o tamanho e forma de estruturas na imagem (HARALICK et al., 1987).

Existem também algoritmos que escolhem o elemento estruturante mais correto em relação as suas necessidades, entretanto, esse procedimento pode ser realizado por um ser humano (SOILLE, 1999).

FACON (1996) define elemento estruturante como: Um conjunto completamente definido e conhecido (forma, tamanho), que é comparado, a partir de uma transformação ao conjunto desconhecido da imagem. A partir dessa transformação é permitido avaliar o conjunto desconhecido. Isso é possível devido ao formato e o tamanho do elemento estruturante permitir testar e quantificar de qual forma o elemento estruturante esta ou não presente na imagem.

Sintetizando, todos os processos e análises da morfologia são dependentes da escolha e também do lugar em que o elemento estruturante se encaixa na imagem.

A partir de elementos estruturantes a morfologia age sobre imagens digitais, que geralmente são definidos em uma malha retangular, conforme a Figura 2.17.

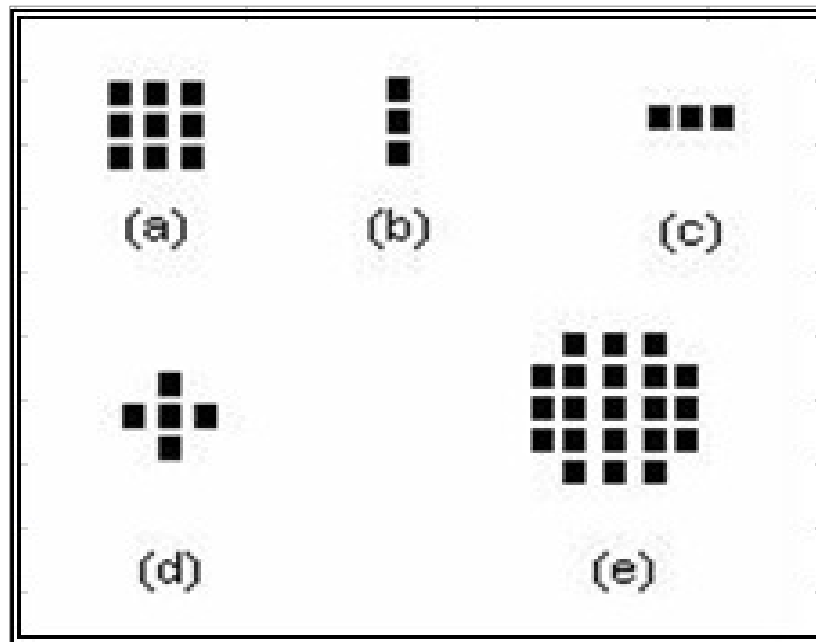


FIGURA 2.17 - EXEMPLOS DE ELEMENTOS ESTRUTURANTES:

(A) Quadrado, (B) Linha Vertical, (C) Linha Horizontal, (D) Cruz e (E) Rhombus.

FONTE: FACON (1996)

Na morfologia matemática, existem algumas operações elementares, que são conhecidos por erosão e a dilatação, que serão descritos a seguir.

2.5.1 Erosão

Para SERRA (1982) a operação de erosão binária pode ser entendida da seguinte maneira: A erosão de um conjunto X pelo elemento estruturante B , segundo a Figura 2.18:

$$ero^B(X) = X \text{ ero } B = \{x \in \varepsilon : B_x \subset X\}$$

FIGURA 2.18 - EROSÃO DE UM CONJUNTO X PELO ELEMENTO ESTRUTURANTE B

FONTE: SERRA (1982).

Alguns efeitos causados em uma imagem pela erosão são: redução das estruturas, os fragmentos das imagens que tem tamanho menor que no elemento estruturante, são removidos, os buracos que existem na imagem são aumentados e os objetos que fazem parte da imagem tendem a ficar mais longe uns dos outros. (SERRA, 1982)

Na Figura 2.19, é demonstrada a imagem original e a imagem resultante da erosão utilizando um elemento estruturante quadrado de tamanho 2, com três iterações.

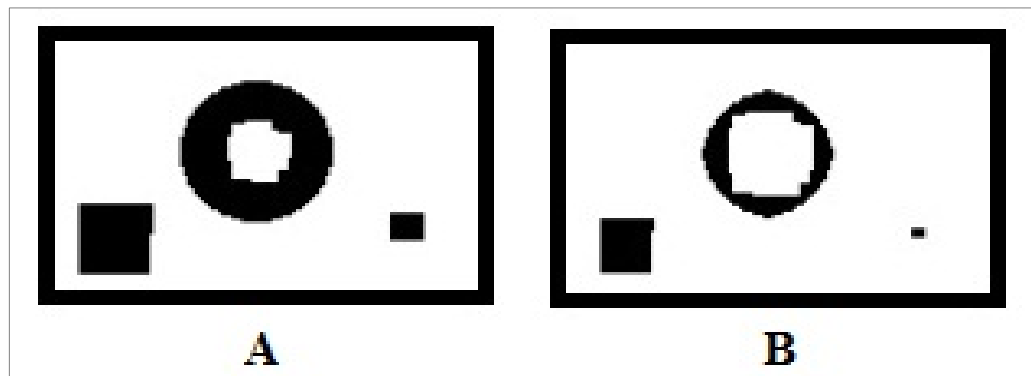


FIGURA 2.19 - EXEMPLO DE EROSÃO

A - Imagem Original, B - Imagem Após Erosão

FONTE: STERNBERG (1986).

Analisando esse exemplo, notamos que o quadrado menor teria desaparecido se fosse aplicado mais uma iteração. O buraco que é visto no meio do círculo, teve seu tamanho aumentado e o quadrado maior teve seu tamanho reduzido em relação ao original. Como observado nos três casos, os objetos ficaram menores em relação a imagem original, esse é o principal resultado obtido fazendo uso da erosão. Com objetos menores o processamento da imagem obtém resultados melhores se comparado com objetos maiores que não se beneficiaram da erosão.

2.5.2 Dilatação

Uma definição para uma operação de dilatação binária segundo SERRA (1982) pode ser dada da seguinte forma: A dilatação de um conjunto X pelo elemento estruturante B é:

$$dil^B(X) = X \text{ dil } B = \{x \in X : B_x \cap X \neq \emptyset\}$$

FIGURA 2.20 - A DILATAÇÃO DE UM CONJUNTO X PELO ELEMENTO ESTRUTURANTE B

FONTE: SERRA (1982).

Alguns efeitos causados em uma imagem pela dilatação são: Aumento dos objetos que pertencem a imagem, preenchimento de buracos pequenos presentes na imagem e união dos objetos mais próximos, como pode ser observado na Figura 2.21 (STERNBERG, 1986).

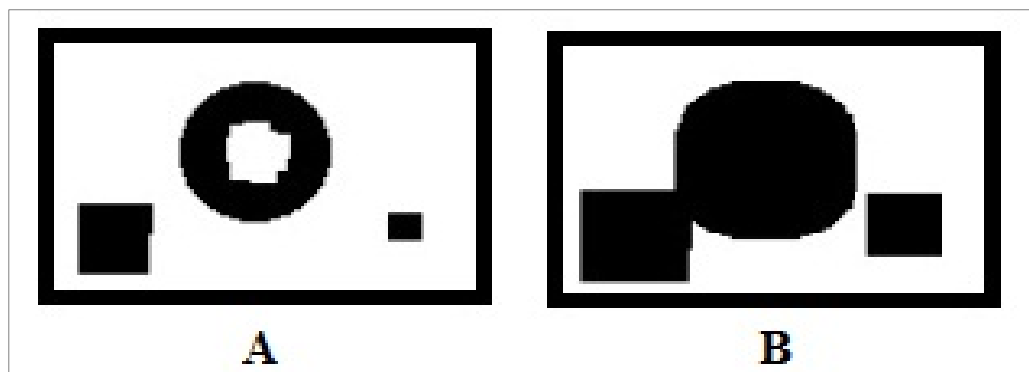


FIGURA 2.21 - EXEMPLO DE DILATAÇÃO

A - Imagem original, B - Imagem após dilatação usando um elemento estruturante quadrado e fazendo uso de cinco iterações.

FONTE: STERNBERG (1986).

É possível notar que os efeitos já citados da dilatação são aplicados nesta imagem. Existia um buraco no meio do círculo que após a dilatação desapareceu, o quadrado maior se uniu ao círculo e por fim o quadrado menor teve seu tamanho

aumentado. Com isso notamos que todos os objetos tiveram seus tamanhos aumentados, o que caracteriza como efeito principal da dilatação.

A erosão e dilatação visam corrigir defeitos na imagem, mas ambas modificam o tamanho dos objetos da imagem, a erosão diminuindo e a dilatação aumentando a imagem. Mas existem duas operações que também corrigem defeitos e eliminam ruídos sem alterar significativamente o tamanho dos objetos, que são conhecidos como abertura e o fechamento (STERNBERG, 1986; SERRA, 1988).

2.5.3 Abertura

Para FACON (1996) a operação de abertura binária utilizando as definições de erosão e dilatação já apresentadas é dada da seguinte maneira:

$$abe^B(X) = X \text{ } abe \text{ } B = (X \text{ } ero \text{ } B) \text{ } dil \text{ } \bar{B} = dil^{\bar{B}}(ero^B(X))$$

FIGURA 2.22 - A ABERTURA DE UM CONJUNTO X PELO ELEMENTO ESTRUTURANTE B.

FONTE: FACON (1996).

A abertura iguala os contornos pelo interior, afasta os objetos, retira partículas que possuem tamanho menor ao elemento estruturante, mas com a grande vantagem de que os elementos ainda presentes na imagem ficam praticamente ou quase idênticos as originais, porém o conjunto aberto possui menos detalhes que o conjunto original (SERRA, 1988).

Na Figura 2.23 pode ser observado o resultado da abertura binária.

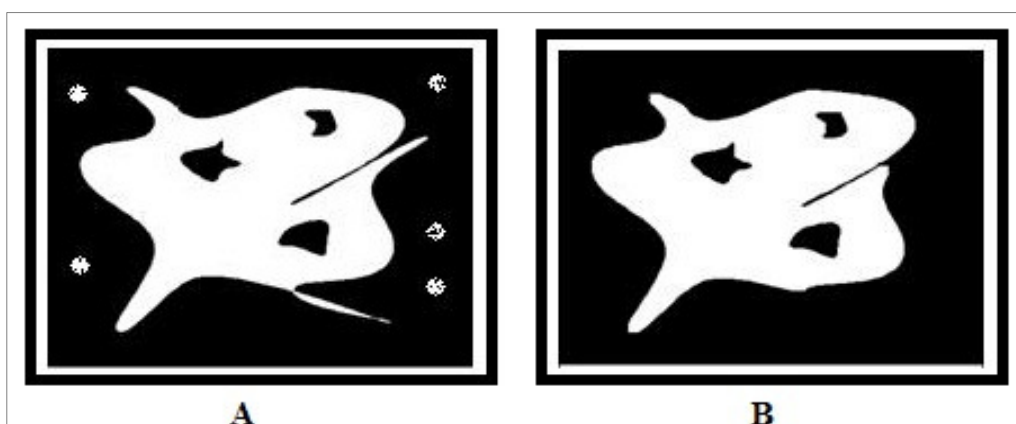


FIGURA 2.23 – RESULTADO ABERTURA BINÁRIA.

A - Imagem original com ruído, B - Imagem após ser aplicada a abertura binária.

FONTE: (SERRA, 1988).

Nota-se que os ruídos existentes na Figura 2.23 A, já não estão mais presentes na Figura 2.23 B, pois foram eliminados com o auxílio do processo de abertura, e a imagem resultante desse processo é bem parecida com a imagem original.

2.5.4 Fechamento

FACON (1996) define a operação de fechamento binário utilizando as definições de erosão e dilatação já apresentadas e dadas da seguinte maneira: O fechamento de um conjunto X pelo elemento estruturante B é:

$$fec^B(X) = X \text{ } fec \text{ } B = (X \text{ } dil \text{ } B) \text{ } ero \text{ } \bar{B} = ero^{\bar{B}}(dil^B(X))$$

FIGURA 2.24 - FECHAMENTO DE UM CONJUNTO X PELO ELEMENTO ESTRUTURANTE B.

FONTE: FACON (1996).

O fechamento suaviza o contorno da imagem pelo exterior, preenche fendas no interior das partículas, juntam as partículas que estão próximas. Os objetos

resultantes após o fechamento tendem a ficar quase idênticos ao modelo original, e como ocorre na abertura, o conjunto fechado possui menos detalhes que o conjunto inicial (SERRA, 1988).

A Figura 2.25 mostra um exemplo do resultado do fechamento binário.

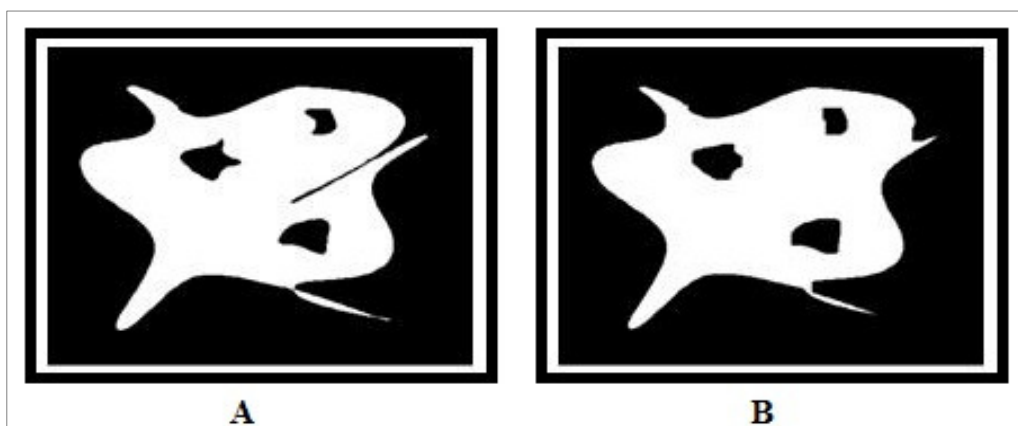


FIGURA 2.25 –RESULTADO FECHAMENTO BINÁRIO

A - Imagem original antes do fechamento, B - Imagem após ser utilizado o processo de fechamento.

FONTE: (SERRA, 1988).

Como pode ser visto na Figura 2.25 B, os efeitos do fechamento suavizaram os contornos dos "furos" presentes no objeto original e eliminou a fenda que existia no interior do objeto.

A abertura e o fechamento diferem da erosão e da dilatação pela propriedade da idempotência, isto é, se for aplicado mais de uma vez, nada muda antes da primeira aplicação. Essa propriedade é crucial no processo de filtragem (FACON, 1996).

2.6 Linguagem C

C é uma linguagem que teve sua origem junto com o advento da teoria de linguagem estruturada e também da criação do computador pessoal (PLAUGER et

al., 1991.) Desta forma, tornou-se com rapidez uma linguagem "popular" entre as pessoas que realizavam programação.

A linguagem C foi utilizada para expandir o sistema operacional UNIX, e hoje é utilizada para desenvolver outras linguagens, como a linguagem C++ e Java (MIZRAHI, 1990; SCHILDT, 1988).

As principais características do C são:

- É uma linguagem de alto nível com uma sintaxe bastante estruturada e flexível, com isso faz com que sua programação seja bastante simplificada;
- Os programas em C são compilados, assim gerando programas executáveis;
- O C compartilha recursos tanto de alto quanto de baixo nível, devido permitir acesso e programação direta do microprocessador. Assim, rotinas cuja dependência do tempo é crítica, podem ser implementadas de maneira fácil usando instruções em Assembly. Devido a isto, o C é a linguagem preferida dos programadores de aplicativos.
- A linguagem C é estruturalmente simples e de grande portabilidade. O compilador C gera códigos mais enxutos e velozes do que muitas outras linguagens.
- Não obstante ser estruturalmente simples, o C não perde funcionalidade porque permite a inclusão de uma farta quantidade de rotinas do usuário. Os fabricantes de compiladores fornecem uma ampla variedade de rotinas pré-compiladas em bibliotecas.

A portabilidade dos programas em C origina-se do uso de bibliotecas de funções para as operações dependentes da máquina, como I/O (entrada e saída) e gerência de memória.

A linguagem C possui poucas palavras-chave, mas um rico conjunto de operadores, inclusive operadores de bits (KERNIGHAN et al., 1986). Por isso, a linguagem C é bastante utilizada em programação de sistemas (compiladores, sistemas operacionais), pois permite manipulação de bits, bytes e os endereços com que o computador funciona basicamente (PLAUGER et al., 1991; SCHILDT, 1988).

2.7 OpenCV

A biblioteca de desenvolvimento OpenCV (*“Open Source Computer Vision Library”*), foi desenvolvida pela Intel e possui mais de 500 funções (WILSON, 2008). É uma biblioteca multiplataforma, de código aberto, totalmente livre ao uso acadêmico e comercial (BRADSKI, 2008).

Foi idealizada com o objetivo de tornar a visão computacional acessível a usuários e programadores em áreas tais como a interação humano-computador em tempo real e a robótica. Muitos dos métodos utilizados para visão computacional já se encontram recriados em C ou C++, assegurando a robustez e a rapidez de processamento.

A resolução do problema torna-se então em encontrar os métodos exatos que permitam uma maior fiabilidade. O programador pode-se concentrar em pesquisar o método mais correto para a resolução desse problema ao invés de concentrar esforços na implementação de algoritmos, centrando o esforço numa análise do seu desempenho e não na programação desses mesmos métodos.

A biblioteca está disponível com o código fonte e os executáveis otimizados para os processadores Intel (WILSON, 2008). Um programa OpenCV, ao ser executado, invoca automaticamente uma DLL (Dynamic Linked Library) que detecta o tipo de processador e carrega, por sua vez, a DLL otimizada para este.

Juntamente com o pacote OpenCV é oferecida a biblioteca IPL (Image Processing Library), da qual a OpenCV depende parcialmente, além da documentação e um conjunto de códigos exemplos (BRADSKI, 2008).

A biblioteca está dividida em cinco grupos de funções:

- Processamento de imagens;
- Análise estrutural;
- Análise de movimento e rastreamento de objetos;
- Reconhecimento de padrões;
- Calibragem de câmara e reconstrução 3D.

Na área de processamento de imagens é importante que as imagens estejam com uma mínima qualidade para que seja possível manipulá-las.

Existem filtros matemáticos com o objetivo de realçá-las de modo que a imagem de saída seja mais adequada à aplicação do que a imagem original.

3 CAPITULO 3

3.1 Materiais

Todo o acervo utilizado nesse trabalho foi cedido pelo Laboratório de Citogenética Animal da UFPR, e foram utilizadas 11 imagens, sendo a maioria no formato JPEG.

As imagens utilizadas nesse trabalho constam no apêndice. A Figura 3.1 mostra algumas dessas imagens.

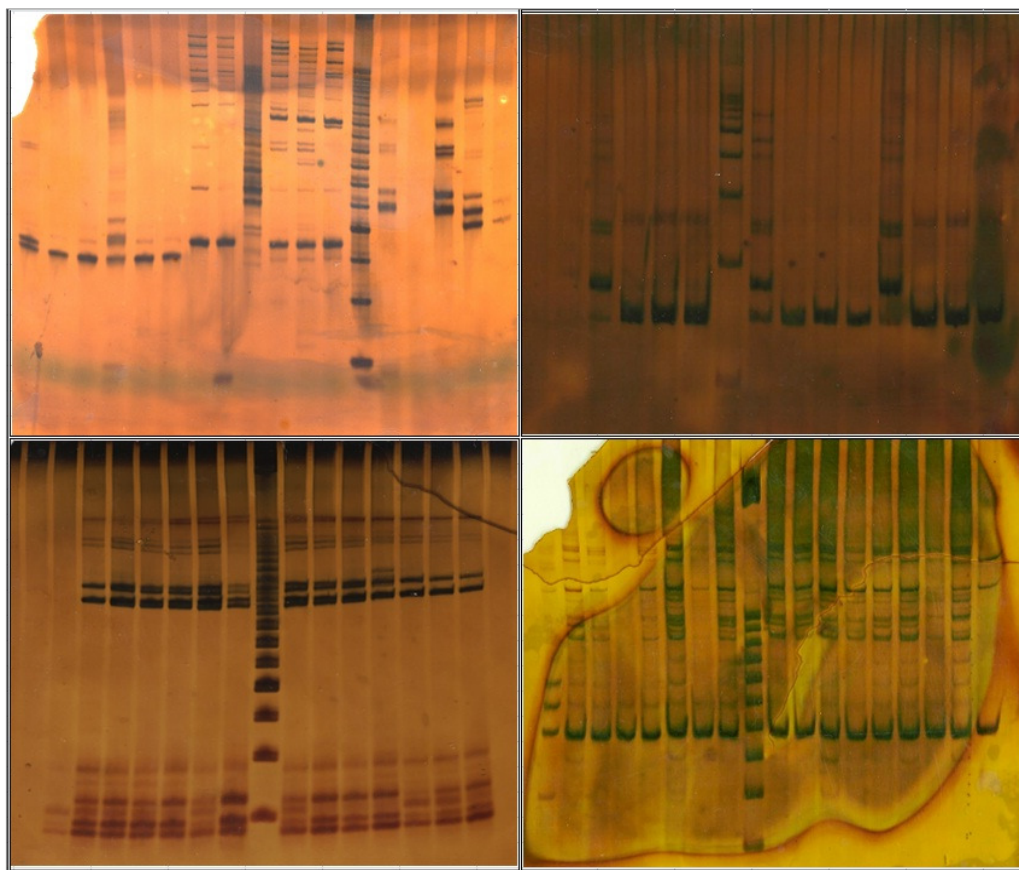


FIGURA 3.1 - ALGUMAS IMAGENS DE GÉIS DE ELETROFORESE UTILIZADAS NESSE TRABALHO.

FONTE: O autor (2011).

3.2 Métodos

3.2.1 Processamento das imagens

- As imagens passaram por um processo de recorte visando eliminar áreas que pudessem prejudicar a execução do algoritmo;
- As imagens originais dos géis foram tratadas com filtros de suavização do tipo Gaussiano e Bilateral para promover o melhoramento das imagens, eliminando ruídos e imperfeições e assim dando mais nitidez às imagens.
- Após a etapa anterior, as imagens foram transformadas em tons de cinza, visando reduzir a quantidade de informações presentes na imagem, dessa forma obtendo os parâmetros corretos para identificar as regiões de interesse;
- Em seguida as imagens foram submetidas a segmentação para que fosse possível o algoritmo reconhecer na imagem às áreas relevantes e também permitir a separação dos objetos em relação ao fundo;
- Para finalizar as imagens foram tratadas através da morfologia matemática binária aplicando-se quando necessário erosão, dilatação, abertura e fechamento que visavam diminuir a quantidade de regiões “brancas”, auxiliando o algoritmo na identificação das colunas. O elemento estruturante utilizado nessa etapa foi do tipo elipse de tamanho 7, e foi realizado 3 iterações, com isso foi possível obter um resultado satisfatório.

3.2.2 Linguagem C

Para o desenvolvimento deste trabalho o algoritmo implementado foi desenvolvido na linguagem C, que é recomendada para o desenvolvimento de aplicações onde existe um balanço entre os dois extremos, o baixo nível e o alto nível. Devido a isto a linguagem C está ficando mais restrita, mesmo tendo se

tornado muito popular, nas aplicações científicas ou aplicações onde não existam linguagens específicas, que auxiliem a vida do programador.

A linguagem C é uma linguagem de alto nível, genérica. Foi criada por programadores para programadores, tendo como objetivos principais, características de flexibilidade e portabilidade (KERNIGHAN et al., 1986).

3.2.3 Biblioteca de desenvolvimento

Como biblioteca de desenvolvimento foi utilizada o OpenCV (*“Open Source Computer Vision Library”*), desenvolvido pela Intel e que possui mais de 500 funções (WILSON, 2008). É uma biblioteca multiplataforma, de código aberto, totalmente livre ao uso acadêmico e comercial (BRADSKI, 2008).

4 CAPITULO 4

4.1 Resultados

Os resultados apresentados a seguir foram validados por um especialista, onde foram feitas inúmeras observações, tais como:

- Qualidade da imagem estava apropriada;
- Verificado cada imagem individualmente;
- Encontrava o marcador corretamente;
- Algoritmo estava encontrando as colunas válidas;
- Demarcando as colunas de forma adequada.

4.1.1 Processamento de cada imagem

A resolução da imagem adquirida contribui diretamente no detalhamento da informação contida. Com uma baixa resolução, a imagem possui menos pontos e torna mais complexa a automatização do processo de detecção de colunas.

Já a alta resolução traz uma riqueza maior nos detalhes (grande quantidade de pontos), deixando a análise mais lenta e exigindo equipamentos mais poderosos.

Com base nessas informações, foram utilizadas imagens com uma resolução de 300 d.p.i. ("*dots per inch*"), que permite um padrão mínimo de qualidade já que os objetos de interesse na imagem são grandes e não precisávamos de detalhes menores.

Cada imagem foi salva em 2 arquivos, sendo um arquivo da imagem original e o outro da imagem pré-processada.

Foi necessária essa separação, pois as imagens originais continham um fundo branco e, em alguns casos, possuíam bordas ou estavam inclinadas, tornando mais complexo o processamento.

Em 7 imagens foram aplicados recortes das regiões que continham as informações. Algumas imagens estavam inclinadas e foram submetidas à rotação para deixar a imagem alinhada corretamente, como apresentado na Figura 4.1.

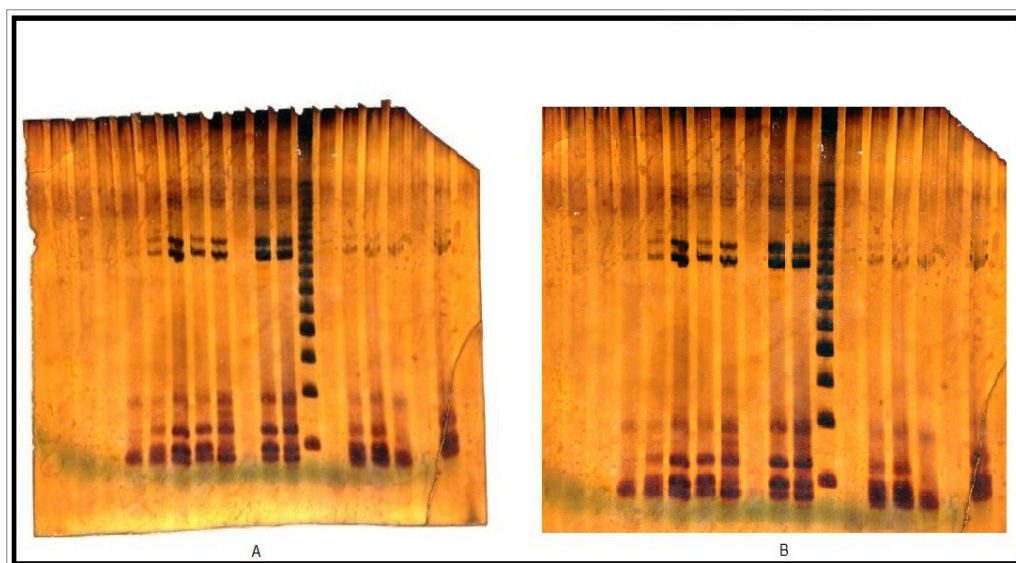


FIGURA 4.1 - IMAGEM APÓS PROCESSO DE ROTAÇÃO

A - Imagem Original, B - Imagem Submetida a Rotação e ao Recorte

FONTE: O autor (2011).

Para obter um nível aceitável de sucesso nos resultados esperados, foram realizados alguns procedimentos nas imagens dos géis de eletroforese, com isso melhorando a qualidade da imagem.

Um conceito importante em processamento de imagens é a Região de Interesse. Entende-se como Região de Interesse a região definida automaticamente a partir de parâmetros obtidos na própria imagem (ou por um usuário) onde o processamento estará totalmente concentrado. Podemos, por exemplo, definir uma região de interesse onde sabemos por antecedência que a iluminação de fundo é constante ou foi corrigida.

Normalmente, nas técnicas de processamento de imagens iremos sempre encontrar problemas no tratamento das bordas da imagem. A criação de regiões de

interesse faz com que apareçam também problemas nas bordas das regiões de interesse. A solução deste problema é completamente dependente do problema ao qual o processamento de imagens está associado, isto é, ao que conhecemos e podemos levar em consideração anteriormente ao processamento.

Com a região de interesse podemos localizar a delimitação de cada coluna presente na imagem do gel, técnica que será utilizada neste trabalho.

No nosso trabalho inicialmente foi aplicado um filtro não-linear, chamado de bi-lateral, visando o realce do contraste e a remoção de ruído, podendo assim dar mais nitidez as imagens e assim evidenciando as regiões que se tinham interesse (Figura 4.2 B).

Após a aplicação do filtro, a imagem foi transformada em tons de cinza, para reduzir a quantidade de informações na imagem, com isso facilitando a obtenção dos parâmetros para identificar as regiões de interesse (Figura 4.2 C).

Feito isto, a imagem foi segmentada (binarizada) pelo *threshold* adaptativo (Figura 4.2 D), para que nas próximas etapas o algoritmo desenvolvido pudesse reconhecer na imagem às áreas relevantes, regiões em branco, e também permitindo a separação dos objetos em relação ao fundo.

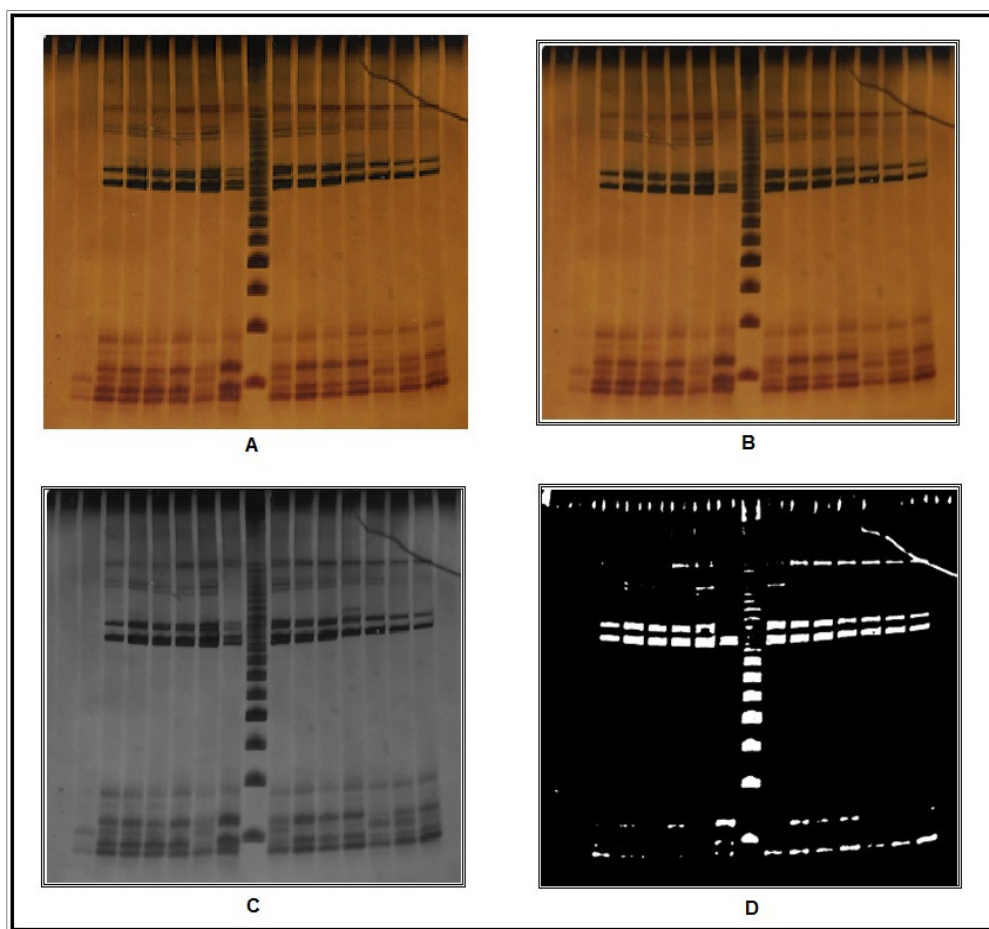


FIGURA 4.2 - RESULTADO DAS ETAPAS DE PRÉ-PROCESSAMENTO.

A - Imagem 5 Original, B- Após Aplicação Do Filtro Bi-Lateral, C - Após Transformação Para Escala De Cinza, D - Após Ser Segmentada Pelo Threshold.

FONTE: O AUTOR (2011).

4.1.1.1 Identificação do marcador

Na etapa para localizar o marcador, o algoritmo desenvolvido, denominado AAGEU, percorre toda a imagem binarizada (branca e preta – Figura 4.2 D) em busca da coluna que possui mais regiões classificadas como banda, que são representadas pela cor branca.

Ao encontrar as coordenadas da região do possível marcador, essas informações são armazenadas.

A partir desse ponto a coluna é percorrida para encontrar a maior sequência de pontos brancos na horizontal, tem início pela esquerda da coluna e vai até sua direita, buscando dois valores, que são os pontos brancos mais distantes da coluna, à esquerda e a direita (Figura 4.3 A).

Ao término desse mapeamento é estabelecida a região lateral onde o marcador tem seu início e seu fim.

Em seguida, a coluna é novamente analisada, agora tendo a finalidade de encontrar o ponto mais alto e mais baixo do marcador (Figura 4.3 B).

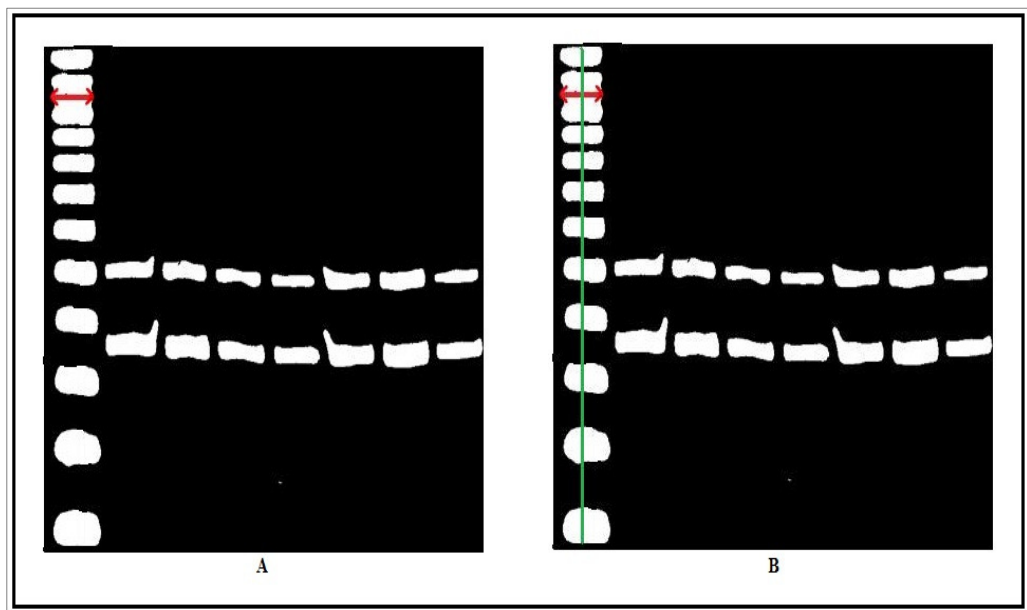


FIGURA 4.3 – ETAPA IDENTIFICAÇÃO DO MARCADOR.

A - Região lateral onde o marcador tem seu início e seu fim, B - Ponto mais alto e mais baixo do marcador.

FONTE: O autor (2011).

Ao final dessa execução obtêm-se as coordenadas que delimitam a região do possível marcador, separando-as do resto da imagem.

4.1.1.2 Identificação das colunas

O objetivo é encontrar os poços e as colunas das imagens, estas regiões serão analisadas em outras etapas, pois contêm as bandas que devem ser processadas e por isso são separadas como regiões válidas.

Esse procedimento auxilia na eliminação de partes não relevantes, fazendo com que seja possível uma classificação apenas nas regiões que se tem maior interesse, ou seja, regiões que realmente possuem material depositado.

A estratégia utilizada foi classificar todas as colunas que possuam 100% de pontos pretos como região de não-coluna.

Entretanto fazendo uso da imagem binarizada pelo "*Threshold*" adaptativo isso não foi possível, pois poucas imagens possuíam colunas totalmente pretas, já que a parte superior da imagem contém áreas escuras ou as colunas eram "tortas".

Dessa forma, foi aplicada uma operação de abertura morfológica na imagem binarizada para diminuir a quantidade de regiões "brancas", o elemento estruturante utilizado nessa etapa foi do tipo elipse de tamanho 7, e foi realizado 3 iterações (Figura 4.4). A imagem resultante foi percorrida para encontrar as colunas cobertas totalmente por pixels pretos. A quantidade de pixels brancos por coluna foi amostrada em um vetor do tamanho da largura da imagem. Em cada posição o vetor armazenou a quantidade de pontos brancos presentes em cada coluna.

O vetor é percorrido e são feitas duas verificações na posição t e em seu vizinho $(t+1)$. A primeira verificação observa se o conteúdo da posição t é igual a zero e se em $t+1$, existe um valor diferente de zero. Se essas condições forem verdadeiras, a coluna t é validada e armazenada como início de um poço, pois possui algum ponto que pode ser de um ou mais spots.

A segunda verificação visa encontrar o final do poço, e é verificado se a posição atual é diferente de zero e a posição ao lado $(t+1)$ é igual a zero, pois indicaria que a sequência de colunas com pontos brancos deixou de existir, e o índice $t+1$ é armazenado e definido como coluna final daquele poço. Essa etapa é executada enquanto t for menor que a largura da imagem.

Com todo esse procedimento, foi possível localizar as coordenadas X de início e fim de todas as colunas que possuem bandas permitindo assim que só as regiões de interesse fossem processadas futuramente.

O procedimento pode ser observado na Figura 4.4.

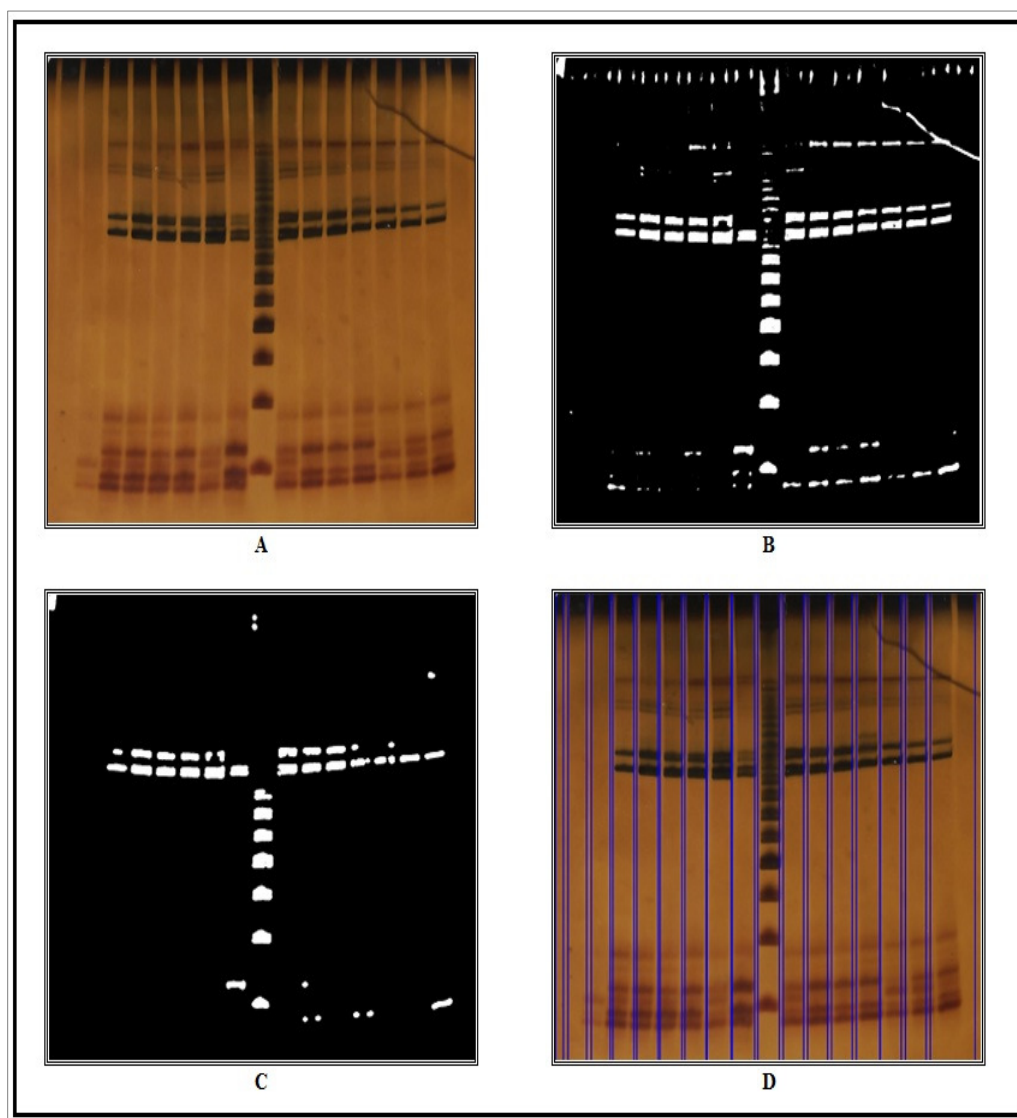


FIGURA 4.4 – ETAPA IDENTIFICAÇÃO DAS COLUNAS

A - imagem com uso do filtro bi-lateral, B - Imagem após procedimento de binarização, C - Imagem após procedimento de abertura morfológica, D - Imagem resultante após execução do algoritmo para encontrar as colunas (marcadas em azul).

FONTE: O autor (2011).

4.1.2 Resultados com o uso do algoritmo para a identificação do marcador

Com base nos testes realizados, foi constatado que em 8 das 11 imagens analisadas ou 73%, o algoritmo desenvolvido para encontrar o marcador teve sucesso.

Nas outras 3 imagens um fator diferente em cada uma influenciou negativamente o resultado final.

No primeiro caso (Figura 4.5 A), o algoritmo não conseguiu percorrer a imagem binarizada a partir da coluna que tem mais pontos brancos para a esquerda, conseqüentemente não encontrou o marcador. Como as primeiras linhas da imagem segmentada (Figura 4.5 B) são preenchidas praticamente só com pixels brancos, isso fez com que o algoritmo entendesse que o começo do marcador estava nas primeiras colunas da imagem.

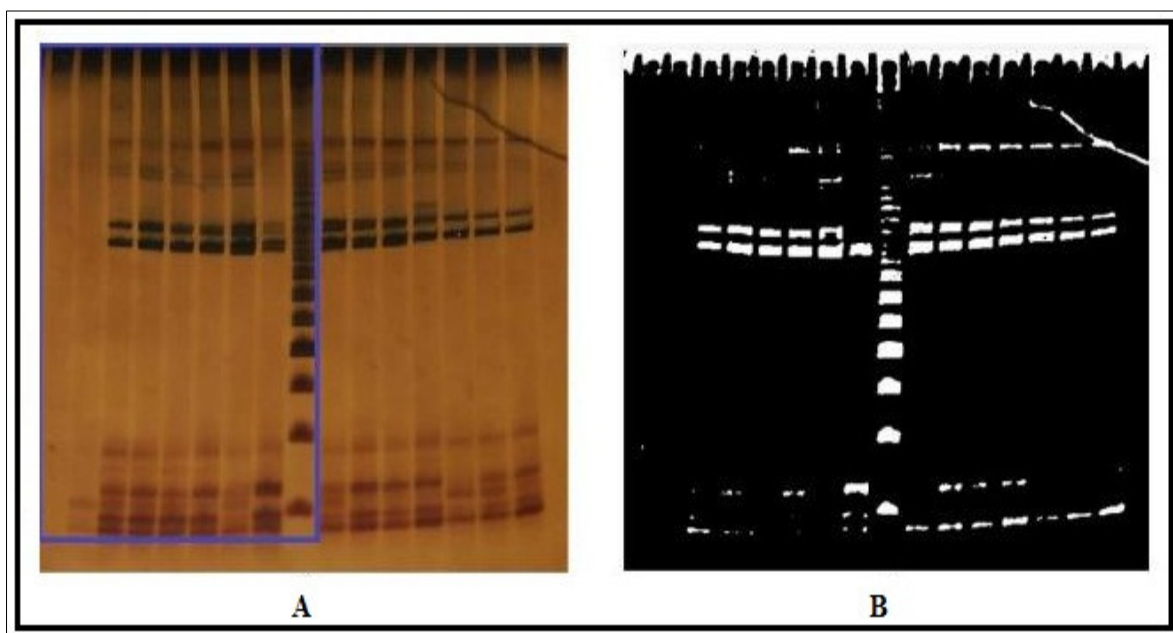


FIGURA 4.5 – PROBLEMA ENCONTRADO NA IDENTIFICAÇÃO DO MARCADOR (CASO 1).

A esquerda está a imagem marcada de forma errada na região do marcador de acordo com o resultado da classificação do algoritmo e a direita sua versão binarizada com o threshold.

FONTE: O autor (2011).

A imagem da Figura 4.6 A, não pode ser classificada de forma correta devido ao seu baixo contraste, deixando a segmentação ineficaz, com isso não foi possível verificar a distinção do que é fundo e objeto.

Como pode ser observado na imagem Figura 4.6 B, foi possível obter apenas algumas regiões que estavam em maior evidência.

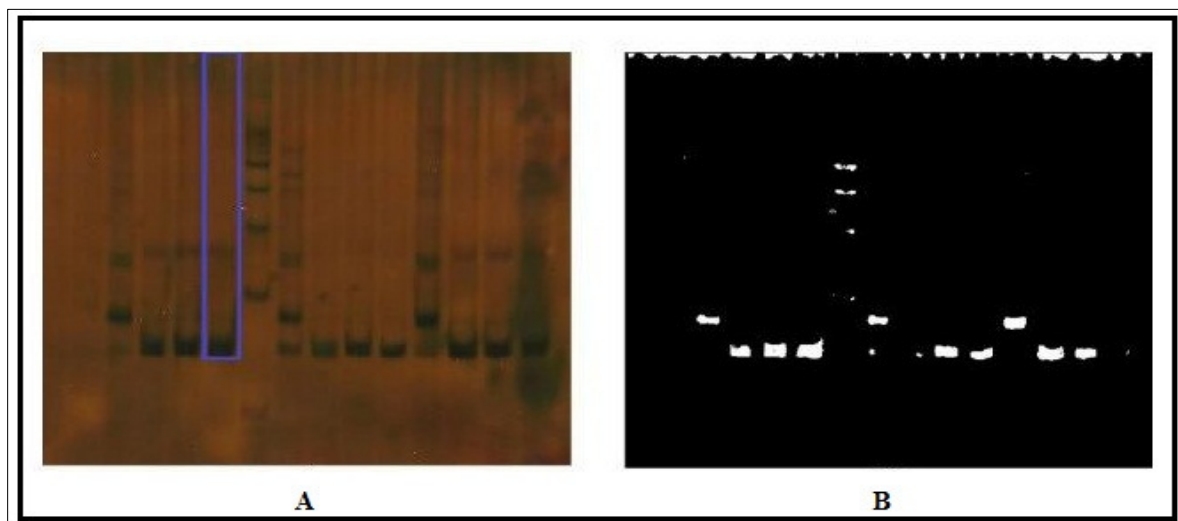


FIGURA 4.6 – PROBLEMA ENCONTRADO NA IDENTIFICAÇÃO DO MARCADOR (CASO 2).

A esquerda está a imagem marcada de forma errada na região do marcador de acordo com o resultado da classificação do algoritmo e a direita sua versão binarizada com o threshold.

FONTE: O autor (2011).

No caso da Figura 4.7 A, a imagem possuía uma grande quantidade de ruído e manchas, por este motivo o algoritmo não concluiu o cálculo da imagem, fazendo com que reconhecesse outra região, que possuía apenas fragmentos da amostra disposta sobre o gel e estava dentro de duas manchas do gel com baixo contraste.

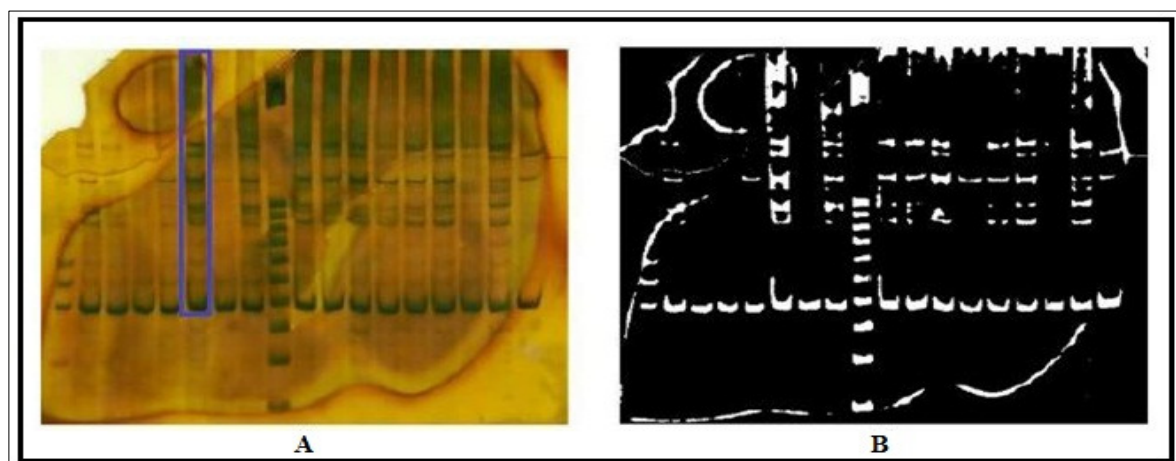


FIGURA 4.7 - PROBLEMA ENCONTRADO NA IDENTIFICAÇÃO DO MARCADOR (CASO 3).

A esquerda está a imagem marcada de forma errada na região do marcador de acordo com o resultado da classificação do algoritmo e a direita sua versão binarizada com o *Threshold*.

FONTE: O autor (2011).

4.1.3 Resultados com o uso do algoritmo para a identificação das colunas

Esta etapa obteve uma taxa de acerto de 82%, levando em consideração o resultado apresentados de cada imagem processada e validada pelo especialista, como pode ser visto na Tabela abaixo.

Tabela 1 – Resultados referentes à identificação das colunas.

Marcação Canaletas			
Imagem	Quantidade Real	Quantidade Encontrada	%
1	18	16	89%
2	17	14	82%
3	18	14	78%
4	8	8	100%
5	17	15	88%
6	14	7	50%
7	17	17	100%
8	17	16	94%
9	16	12	75%
10	13	7	54%
11	18	16	89%
		Total Acertos	82%

FONTE: O autor (2011).

Observando a Tabela 1, nas imagens da Figura 4.8, e 4.9 é possível notar que o algoritmo conseguiu encontrar todas ou pelo menos a maior parte das colunas que existiam nas imagens. O emprego do algoritmo teve grande sucesso, pois os poços estavam alinhados, as imagens tinham boa qualidade evidenciando suas informações.

O algoritmo conseguiu encontrar a coluna inicial e final, onde se encontrava a sequência de pontos brancos, de forma bastante simples, devido às regiões estarem preservadas.

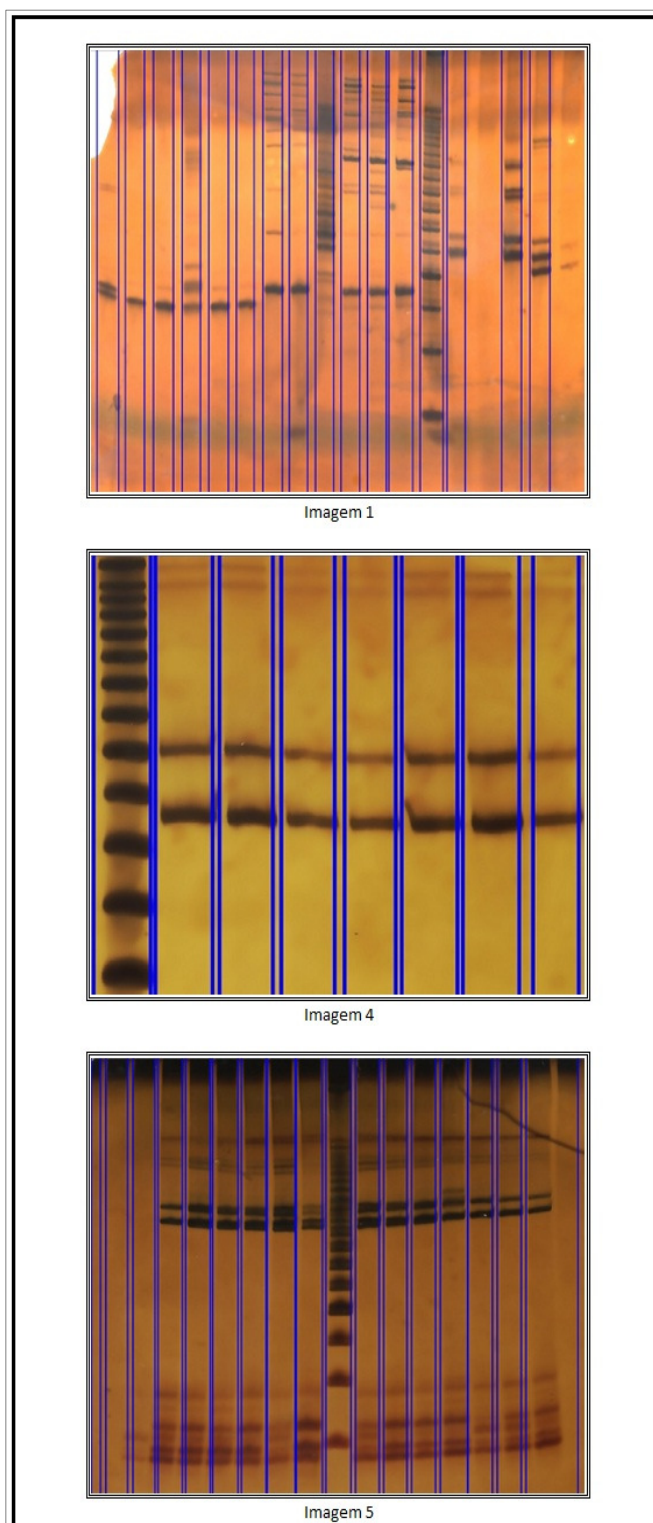


FIGURA 4.8 – RESULTADOS ACIMA DE 85% OBTIDOS NA IDENTIFICAÇÃO DAS COLUNAS

FONTE: O autor (2011).

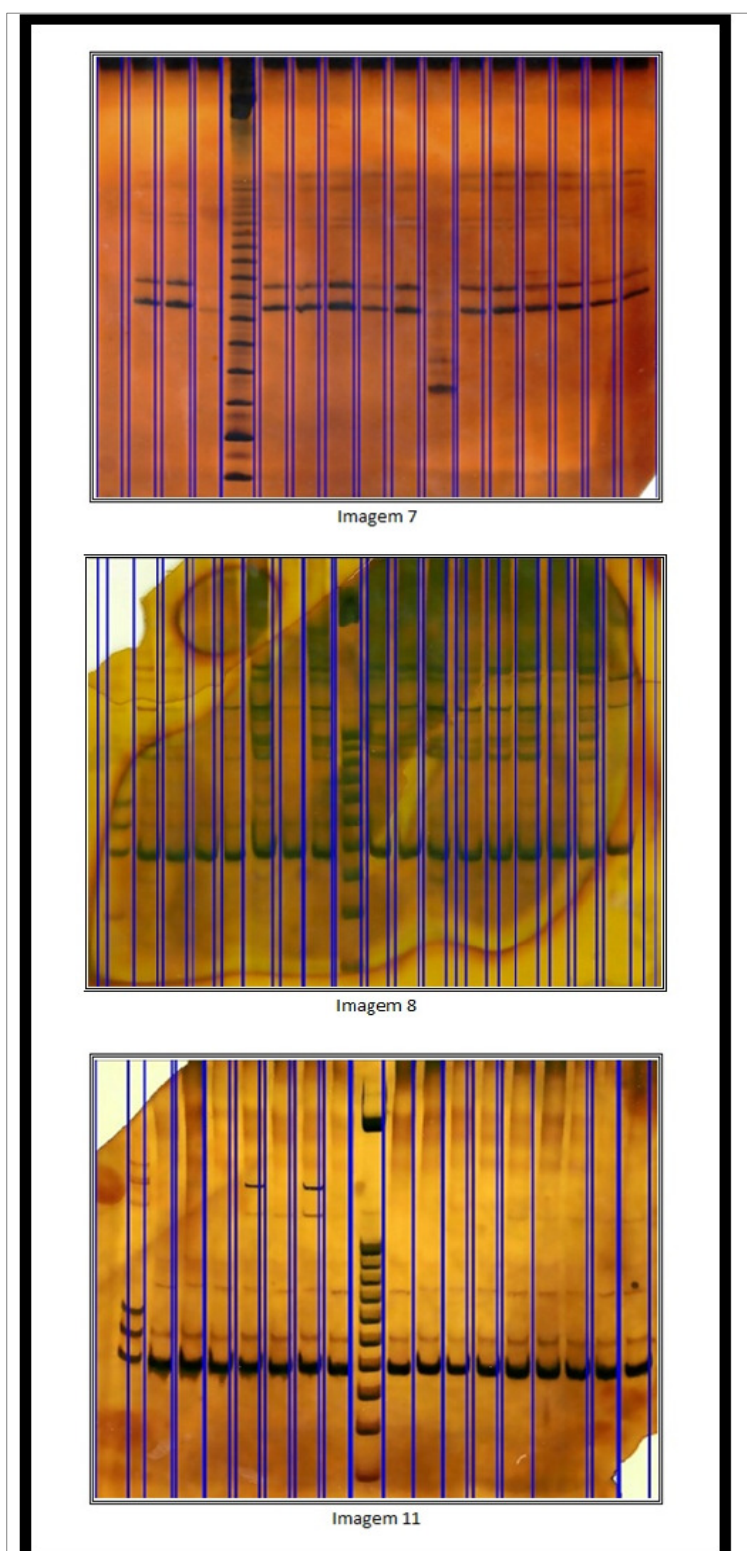


FIGURA 4.9 - RESULTADOS ACIMA DE 85% OBTIDOS NA IDENTIFICAÇÃO DAS COLUNAS

FONTE: O autor (2011).

No caso das imagens da Figura 4.10, surgiram dificuldades para encontrar as colunas de interesse. Essas imagens sofreram interferências devido à quebra do gel durante seu manuseio antes de ser digitalizado, o que influenciou diretamente no processo de segmentação, porém os resultados foram superiores a 75% e inferiores a 85% nos acertos das colunas.

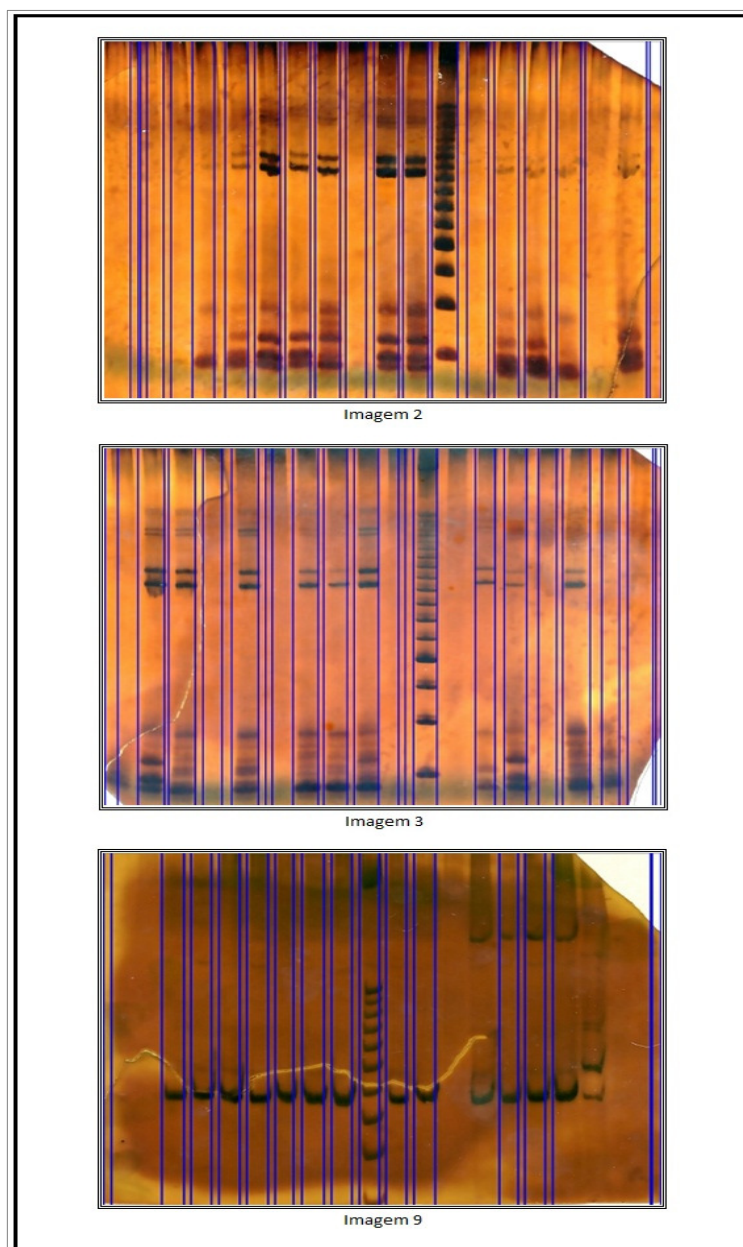


FIGURA 4.10 - RESULTADOS ACIMA DE 75% OBTIDOS NA IDENTIFICAÇÃO DAS COLUNAS

FONTE: O AUTOR (2011).

Devido à baixa qualidade e o fato da estrutura dos poços estarem tortas em diferentes regiões, nas imagens da Figura 4.11, a análise do algoritmo ficou comprometida. Foram encontradas colunas em regiões onde não existiam, ou locais que possuem apenas um poço são detectados como sendo dois, porém ainda assim os resultados alcançaram 50% dos acertos.

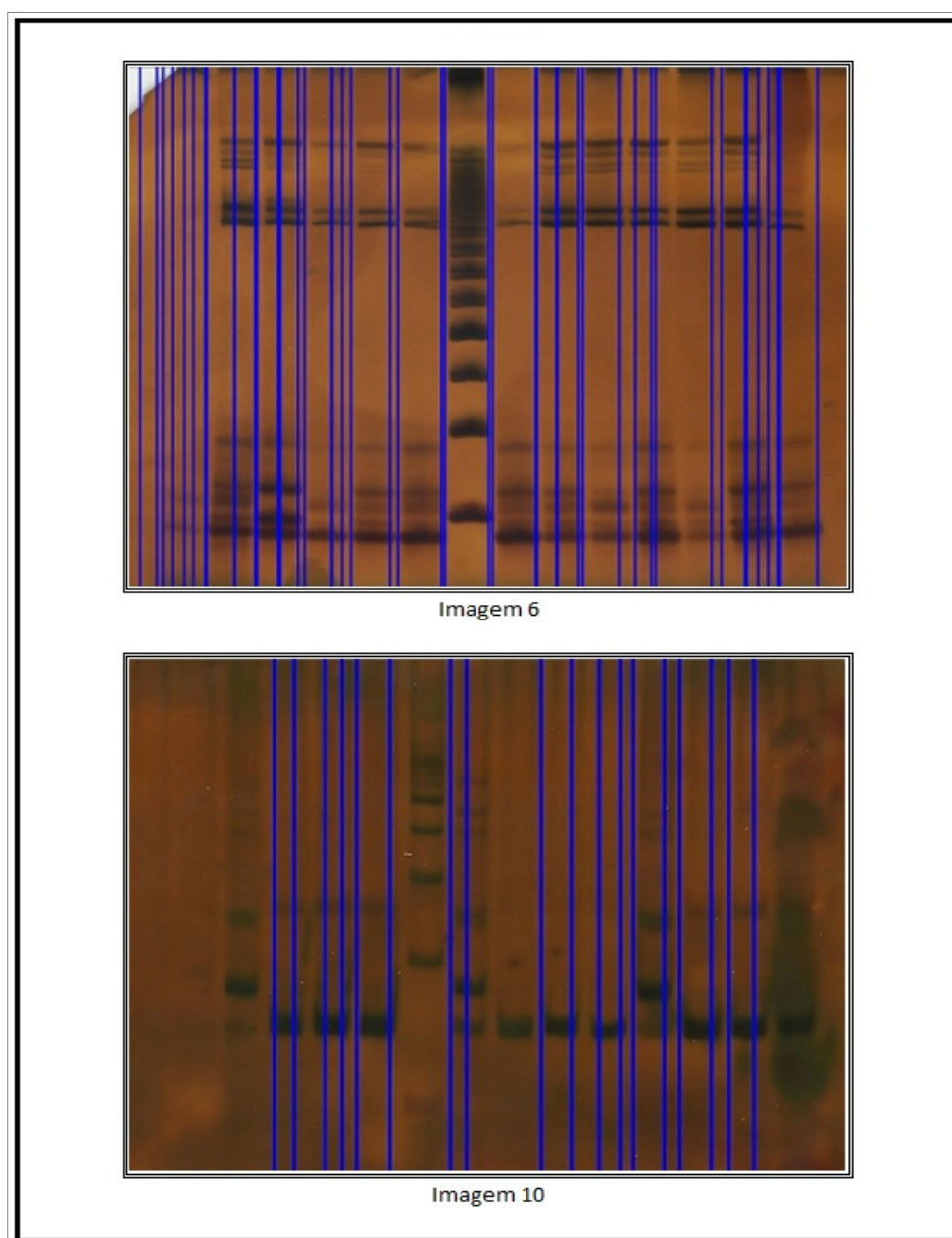


FIGURA 4.11 - RESULTADOS ENTRE 50% E 75% OBTIDOS NA IDENTIFICAÇÃO DAS COLUNAS

FONTE: O autor (2011).

4.2 Comparação\Validação com outros softwares

A seguir teremos alguns dados referentes a comparação e validação da metodologia proposta com dois softwares que se propõem a realizar praticamente as mesmas análises que o algoritmo desta dissertação.

O Projeto SABIA, apresenta uma abordagem computacional diferenciada para a análise de géis de eletroforese unidimensionais em poliacrilamida e agarose, utilizando Inteligência Artificial. É um sistema computacional de acesso público, capaz de ler, interpretar e analisar automaticamente resultados obtidos via imagens digitalizadas de géis de eletroforese é apresentado como alternativa às atuais soluções encontradas no mercado. Este por sua vez tem sua distribuição gratuita e qualquer pesquisador interessado em realizar análise de géis de eletroforese pode fazer o uso dessa ferramenta, sem ter custo com a compra de licenças (SOARES, 2010).

Já o software TotalLab Quant é uma base ampla e de alta funcionalidade para análise de imagem. Ele está estruturado em torno das aplicações mais comuns da ciência da vida, tornando-se uma grande ferramenta para uma variedade de pesquisas. O módulo de Análise do TotalLab Quant 1D é perfeito para uma gama de funções, a partir da determinação do tamanho molecular é possível quantificar com precisão o material da banda. É possível realizar uma análise totalmente automática, uma análise manual, ou uma combinação dos dois. Este software é de uso proprietário, ou seja, caso o pesquisador queira fazer o uso desses recursos é necessário fazer investimentos na compra de licença de uso, o que acaba aumentando o custo e inviabilizando o uso nos laboratórios de pesquisa (NONLINEAR DYNAMICS, London, UK).

No caso do marcador, foi possível realizar a comparação somente com o software SABIA, pois o aplicativo TotalLab Quant não traz em sua interface a busca somente do marcador.

Observando a Tabela 2, é possível notar que o software SABIA, conseguiu encontrar 10 marcadores. Mas algumas observações devem ser feitas, realizando a comparação foi possível notar que nas imagens 7, 9, 10 e 11, o software encontrou

o marcador, mas fez sua marcação fora do limite da coluna estipulada como marcador. Isto não pode ser considerado um erro grave, mas o algoritmo desenvolvido não possui esse tipo de problema, nas imagens que conseguiu encontrar o marcador, fez sua delimitação do tamanho adequado da coluna.

Tabela 2 - Comparação com o software SABIA, para encontrar os marcadores.

	Imagem	1	2	3	4	5	6	7	8	9	10	11
Encontrou	AAGEU	SIM	SIM	SIM	SIM	NAO	SIM	SIM	NAO	SIM	NAO	SIM
Marcador	SABIA	NÃO	SIM	SIM	SIM	SIM	SIM	SIM*	SIM	SIM*	SIM*	SIM*

*Encontrou o marcador, mas fez a marcação das mesmas ultrapassando o limite da canaleta.

FONTE: O autor (2011).

Na figura 4.12, pode ser observado dois casos. No primeiro a imagem 1, onde o software proposto conseguiu encontrar adequadamente o marcador, e o SABIA teve dificuldade, demarcando duas colunas como sendo o marcador.

No segundo caso temos a imagem 5, que na metodologia proposta não obteve sucesso, e com o uso do SABIA, foi possível encontrar o marcador.

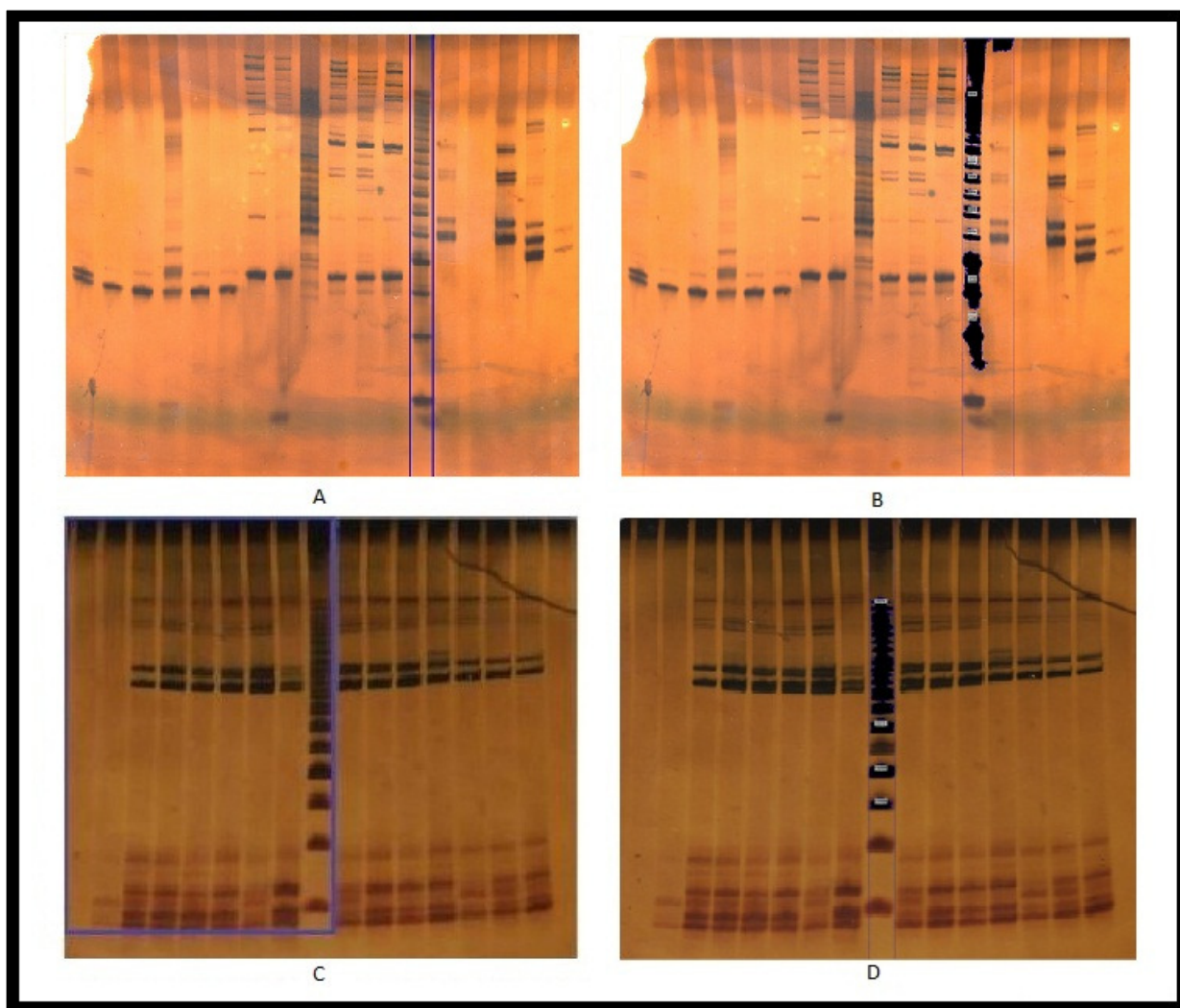


FIGURA 4.12 – RESULTADOS OBTIDOS NA COMPARAÇÃO DA IDENTIFICAÇÃO DO MARCADOR COM O SOFTWARE SABIA.

A - Resultado obtido com o software proposto na imagem 1; B - resultado obtido com o SABIA; C - resultado obtido com o software proposto na imagem 5; D - resultado obtido com o SABIA.

FONTE: O autor (2011).

Para realizar a comparação das colunas encontradas, foram utilizados os dois softwares já citados, pois ambos realizam este tipo de procedimento durante seu processamento.

Tabela 3 - Comparação com o AAGEU, SABIA e o TotalLab Quant, para encontrar as colunas.

	Imagem	Quantidade Real	Quantidade Encontrada		
			AAGEU	SABIA	Total lab Quant
Marcação Canaletas	1	18	16	14	18
	2	17	14	7	12
	3	18	14	4	17
	4	8	8	5	8
	5	17	15	14	14
	6	14	7	12	12
	7	17	17	11	16
	8	17	16	16	16
	9	16	12	15	0
	10	13	7	11	2
	11	18	16	18	0
Total Canaletas		173	142	127	115

FONTE: O autor (2011).

Pode-se notar que os resultados obtidos em relação ao software SABIA, são quase os mesmos quanto a números. Porém, o algoritmo AAGEU traz resultados mais otimizados, como podem ser visto nas imagens do resultado do SABIA. Nota-se que o processamento das imagens com o SABIA, traz alguns erros, tais como:

- Marcações de colunas onde não existem;
- Marcação de duas colunas onde existe apenas uma;
- Demarcação inadequada das colunas.

Em comparação com o software TotalLab Quant, na maioria das imagens ele consegue encontrar perfeitamente todas as colunas existentes. Somente em duas imagens o software não conseguiu encontrar nenhuma coluna corretamente, demarcando mais de uma coluna, onde existia uma coluna válida somente.

Mas também foram observados alguns problemas, como:

- Demarcação de colunas de forma errada, pois esse software faz uso de uma métrica onde ele encontra uma coluna e as demais seguem um padrão de distância uma da outra, fazendo com que algumas colunas não fiquem enquadradas corretamente;
- Marcações de colunas onde não existem;

- Marcação de duas ou mais colunas, onde existem apenas uma.
- Marcação de colunas tortas ou de forma não alinhadas.

Na imagem 4.13, é possível notar que a imagem 5, fazendo uso dos três softwares comparados, os três, apresentam um resultado adequado ao resultado esperado por um especialista.

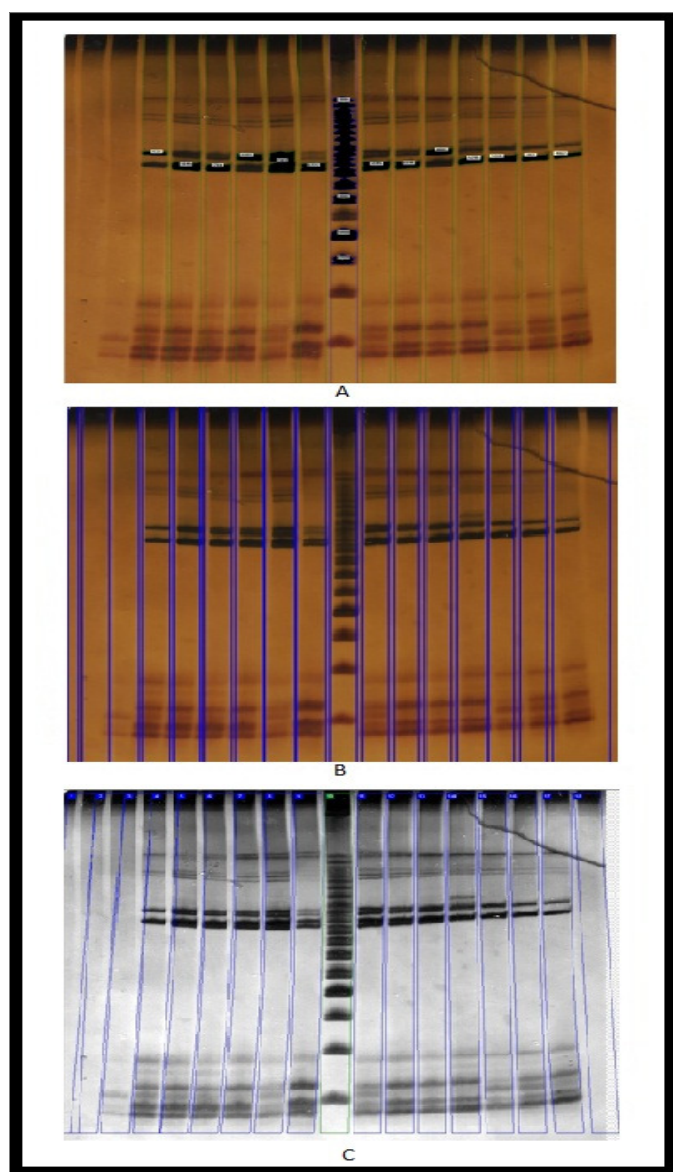


FIGURA 4.13 – RESULTADO DA COMPARAÇÃO DA IMAGEM 5.

A - Resultado obtido com o SABIA na imagem 5; B - resultado obtido com o AAGEU;
C - resultado obtido com o software TotalLab Quant.

FONTE: O autor (2011).

Na imagem 4.14, traz a imagem 9, e nota-se alguns problemas já citados, como a marcação de colunas onde não existem e a marcação de duas ou três colunas onde existe apenas uma.

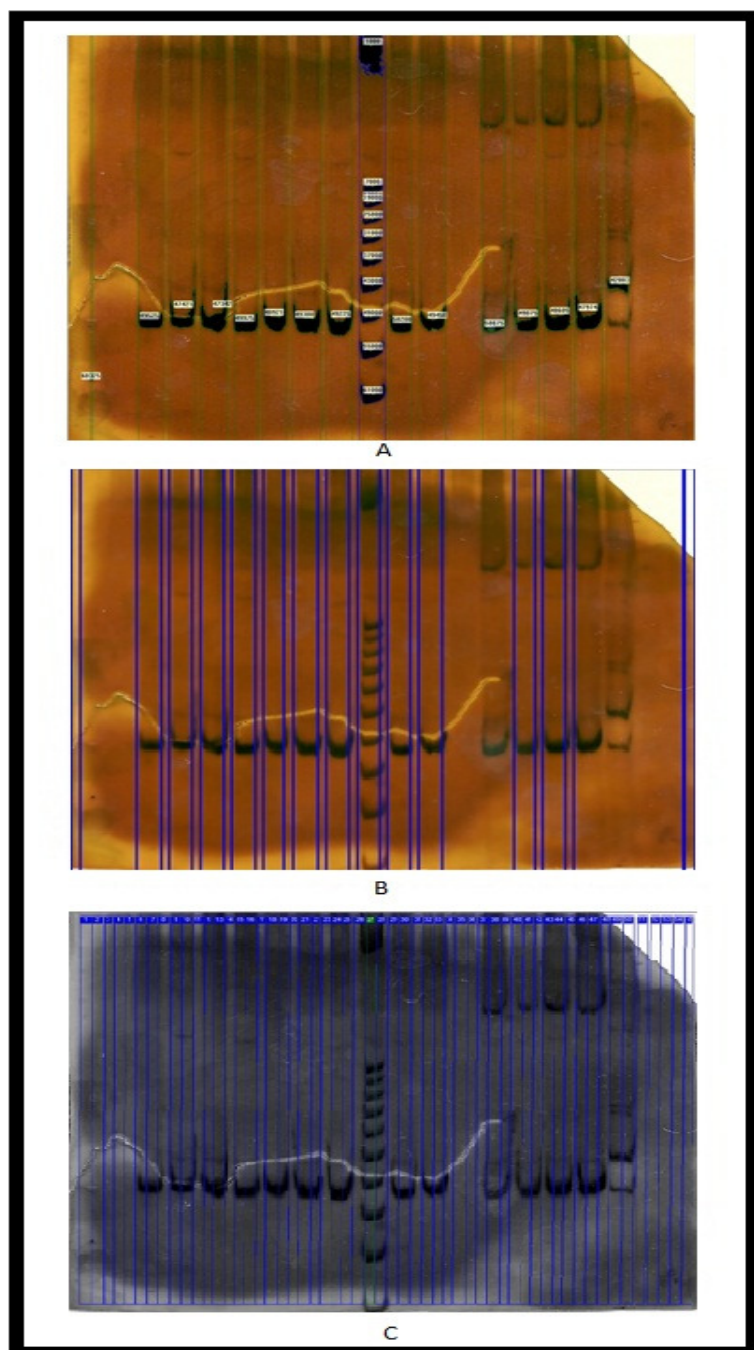


FIGURA 4.14 - RESULTADO DA COMPARAÇÃO DA IMAGEM 9.

A - Resultado obtido com o SABIA na imagem 9; B - resultado obtido com o AAGEU;
C - resultado obtido com o software TotalLab Quant.

FONTE: O autor (2011).

Já a imagem 4.15, traz a imagem 10, e observam-se outros problemas já citados, como a demarcação inadequadas de colunas, marcação de colunas tortas ou de forma não alinhadas.

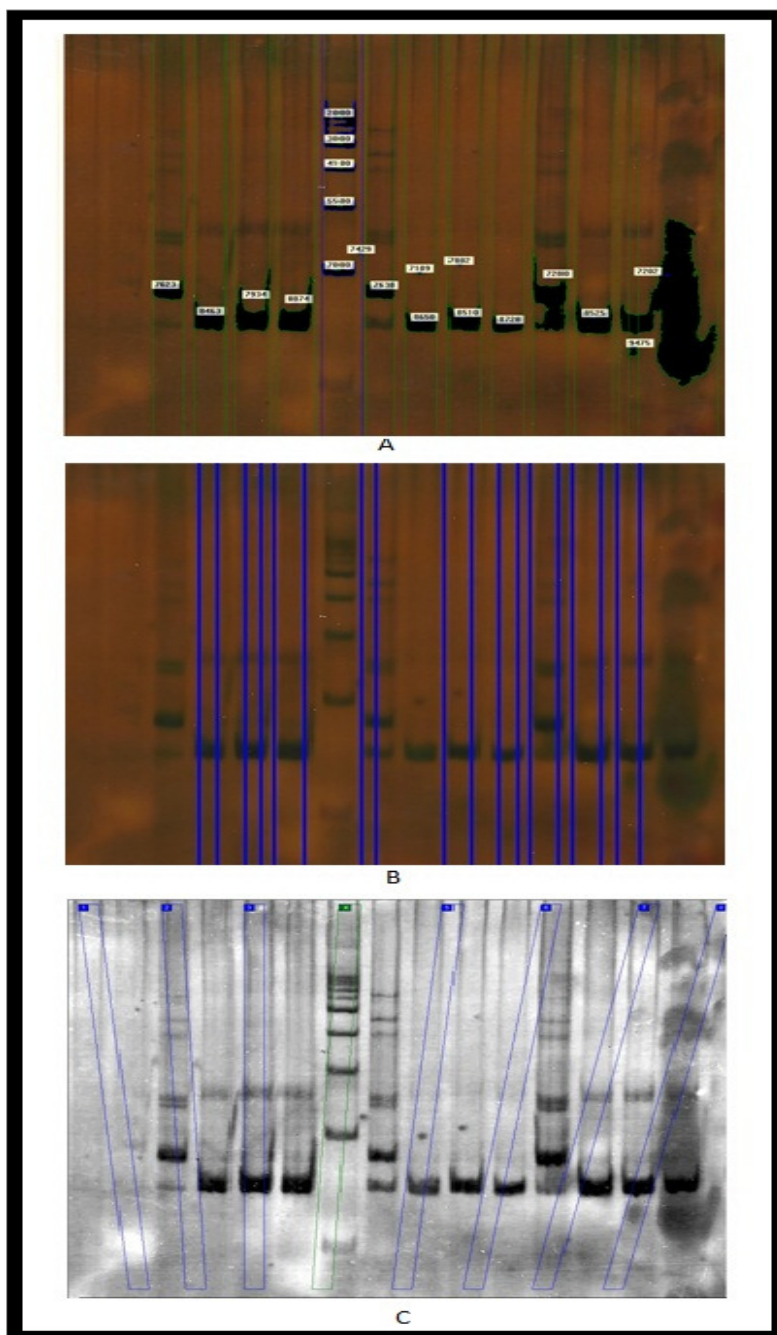


FIGURA 4.15 - RESULTADO DA COMPARAÇÃO DA IMAGEM 10.

A - Resultado obtido com o SABIA na imagem 10; B - resultado obtido com o AAGEU; C - resultado obtido com o software TotalLab Quant.

FONTE: O autor (2011).

5 CAPITULO 5

5.1 Discussão

A análise manual de imagens de eletroforese é um processo subjetivo e tedioso para os pesquisadores, a análise automatizada de imagens de eletroforese com o uso da informática, mais precisamente com o uso de técnicas de processamento de imagens é interessante porque reduz erros humanos e torna o processo de interpretação mais rápido, podendo inclusive ser utilizada para estudar grandes bancos de dados de amostras. Portanto, uma ferramenta capaz de realizar a análise automatizada de forma precisa e eficiente seria de grande valia para as áreas de Biologia Molecular e Genética, assim poderia reduzir o tempo das pesquisas e seus custos.

A escolha da linguagem C para o desenvolvimento do algoritmo foi devido ser estruturalmente simples e de grande portabilidade. O compilador C gera códigos mais enxutos e velozes do que muitas outras linguagens e também não perde funcionalidade porque permite a inclusão de uma farta quantidade de rotinas do usuário.

Foi utilizado o processamento de imagens visando melhorar o aspecto visual de certas feições estruturais para o analista humano, eliminar ruídos presentes na imagem que podiam dificultar a análise e fornecer outros subsídios para a sua interpretação, com isso gerando imagens com melhor qualidade e obtendo resultados satisfatórios.

A delimitação do escopo para a identificação das colunas e do marcador foi estabelecida por ser a base principal para a realização de outros trabalhos envolvendo a análise de géis de eletroforese. Como por exemplo, a partir dessa delimitação é possível aperfeiçoar o algoritmo, visando agora a busca pelas bandas e posteriormente realizando a sua quantificação com base nas informações do marcador.

Durante a execução deste trabalho ficou evidente a necessidade do aprimoramento do manuseio dos géis, pois a qualidade da aquisição do gel (sem manchas, sem rachaduras, etc.) e o tipo de gel interferem diretamente no processamento digital das imagens.

Os géis utilizados na aquisição do nosso acervo possuíam uma consistência mole, fazendo com que durante a fase de aplicação da corrente elétrica os fragmentos se deslocassem sobre a superfície causando deformações no material e gerando o chamado efeito sorriso na parte inferior.

Entretanto, onde as informações estavam presentes e preservadas, a metodologia proposta foi capaz de detectar o posicionamento correto das colunas.

Merece destaque o fato que o algoritmo na maioria das vezes encontrou as colunas onde realmente existia material, e o no fato de fazer a delimitação adequada do tamanho da coluna, os softwares comerciais existentes que realizam esse trabalho, não delimitam as posições de forma precisa das colunas, isso pode afetar no resultado final da análise.

Com imagens melhores a análise se torna menos complexa e permite que as ferramentas desenvolvidas auxiliem os pesquisadores de uma forma mais eficaz.

Em relação à comparação com os outros softwares, podemos dizer que o AAGEU cumpre uma primeira etapa que é a de localizar o marcador e posteriormente identificar as colunas presentes na imagem de eletroforese. Mas merecem destaque algumas dificuldades que até mesmo softwares comerciais possuem, visto que são desenvolvidos para atender uma demanda grande e a profissionais exigentes, que elencamos a seguir:

- Marcação de colunas onde não existem;
- Marcação de duas colunas onde existe apenas uma;
- Demarcação inadequada das colunas, sendo elas de forma torta ou não alinhadas corretamente;
- Uso de métricas para encontrar as colunas presentes na imagem, como mostrado anteriormente essa técnica acaba fazendo com que algumas colunas não fiquem enquadradas corretamente.

Contudo, com o aprimoramento da metodologia desenvolvida, uso de outras técnicas envolvendo processamento de imagens, fazendo o uso de imagens

com qualidade e forma de aquisições melhoradas, a maioria desses problemas podem ser resolvidos obtendo assim um maior grau de precisão. Também é importante que sejam testados outros géis para encontrar a solução mais eficiente do procedimento.

5.2 Conclusões

Com este trabalho foi possível demonstrar que a detecção do marcador e a marcação das colunas são viáveis fazendo uso de um método simples e de rápida execução que está baseada em alguns procedimentos de aprimoramento da imagem, como:

- Recortes das imagens;
- Aplicação de filtros de suavização;
- Transformação das imagens em tons de cinza;
- Na binarização das imagens;
- Utilização de operações morfológicas;
- Contagem de valores brancos ou pretos;
- Entre outros.

Ainda existem algumas limitações envolvendo a qualidade da imagem adquirida, pois como mostrado nos resultados não existe um padrão nas cores e existem muitos fatores, tais como:

- Técnica e corante Utilizados;
- Tempo das amostras e forma de armazenamento;
- Deformação do gel que dificultam a detecção das colunas e/ou dos marcadores;
- Contraste da imagem;
- Manchas no gel;
- Rachaduras no gel;
- Etc.

Como projetos futuros a este trabalho, elencamos:

- Possibilidade de análise de géis de eletroforese que apresentem duas dimensões de dados (géis 2D), como no caso de géis de proteomas;
- Conversão da atual solução computacional de busca dos artefatos para bibliotecas de acesso dinâmico que possibilitem seu acesso a partir de vários outros sistemas operacionais;
- Implementação de funções para melhora automatizada da imagens, retirando ruídos e rachaduras nos géis Desenvolver uma nova estratégia para realizar a quantificação das bandas, ou seja, definir seu peso molecular.

O último item tem forte apelo no meio de pesquisa, pois sobrecarrega as atividades dos pesquisadores, que muitas vezes realizam anotações e cálculos manualmente, com o desenvolvimento dessa estratégia seria otimizado tempo e trabalho.

Softwares de domínio público que fazem esse trabalho são escassos, com isso os pesquisadores ficam atrelados as empresas comerciais, na aquisição de licenças e isso aumenta muito o custo nos laboratórios de pesquisa.

REFERÊNCIAS

ANDREWS, H., TESCHER, A., AND KRUGER, R. **Image processing by digital computer**. IEEE Spectrum, 9(7):20–32. 1972.

AYALA, F.J. **The mechanisms of evolution**. In: **EVOLUTION**. San Francisco, Scientific American, 1978. p.14-27.

AZEVEDO, M. DE O. **Técnicas básicas em biologia molecular**. Brasília, DF: Editora da UnB, 2003 211p.

BANON, G., BARRERA, J., AND BRAGA-NETO, U., Eds. **Mathematical Morphology and its Applications to Signal and Image Processing**. INPE, São José dos Campos, 2007. Proceedings of the 8th International Symposium on Mathematical Morphology.

BLOCK, RICHARD J. (JOSEPH), 1906-; DURRUM, S. EMMET; ZWEIA, GUNTER. **A manual of paper chromatography and paper electrophoresis**. 2. ed. rev. New York: Academic Press, 1958. 710 p.

BRADSKI, G. E KAEHLER, A., **Learning OpenCV**, O'Reilly, 2008.

CARVALHO, ADRIANO ANDREI DE. **Fundamentação Teórica para Processamento Digital de Imagens**. Departamento de Ciência da Computação - Universidade Federal de Lavras - Lavras - MG, 2003.

CASTLEMAN, KENNETH R. **Digital Image Processing**, Prentice-Hall, 1995.

COLTUC, D., BOLON, P., AND CHASSERY, J.-M. **Exact histogram specification**. IEEE Transactions on Image Processing, 15(5):1143–1152. 2006.

CONCI A, AZEVEDO E: **Computação Gráfica - Teoria e Prática**. Rio de Janeiro: Prentice Hall PTR 2007.

DOUGHERTY, E., AND LOTUFO, R. **Hands-on Morphological Image Processing**. SPIE Publications, Bellingham, WA, 2003.

DURAND, F. AND DORSEY, J. **Fast bilateral filtering for the display of high-dynamic-range images**, Proc. 29th ACM SIGGRAPH Conf. on Computer Graphics, San Antonio, USA, 2002.

EISEN, J.A. Mechanistic basis for microsatellite instability. In: Goldstein, D.B.; Schlötterer, C. (eds), **Microsatellites - evolution and applications**, Oxford University Press, Oxford, UK, cap. 4, p.34-48, 1999.

ELAD, MICHAEL. **On the Origin of the Bilateral Filter and Ways to Improve It**. IEEE Transactions On Image Processing, Vol. 11, No. 10, October 2002.

ENGEL, S.R.; LINN, R.A.; TAYLOR, J.F.; DAVIS, S.K. Conservation of microsatellite locos across species of artiodactyls: implications for population studies. **Journal of Mammalogy**, v.77, n.2, p.504-518, 1996.

FACON, J. **Morfologia Matemática: teoria e exemplos**. Curitiba: Universitária Champagnat da Pontifícia Universidade Católica do Paraná, 1996. 320 p.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores em análise genética**. Brasília: Embrapa – Cenargem, 3^a ed., 220p., 1998.

FILHO OM, NETO HV: **Processamento Digital de Imagens**. Rio de Janeiro: Brasport 1999.

GONELA, A. **Aplicação de marcadores microssatélites de *Sus scrofa domestica* caracterização genética de populações de *Sus scrofa* sp (porco-Monteiro) e *Tayassu pecari* (queixada)**. Ribeirão Preto, 2003. 88p. Tese (Doutorado em Genética) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo.

GONZALES, R.C.; WOODS, R.E. **Processamento de Imagens Digitais**. São Paulo: Edgard Blücher Ltda, 2000. 509p.

HANCOCK, JM. **Microssatélites e outras seqüências simples: contexto de genômica e mecanismos de mutação. Em microssatélites: Evolução e aplicações**. Editado por Goldstein D, C. Schlötterer New York, Oxford University Press; 1999:1-9.

HANCOCK, LES E KRIEGER, MORRIS. Manual **de Linguagem C**, Rio de Janeiro, Campus, 1985. 182p.

HARALICK, R., STERNBERG, S., AND ZHUANG, X. Image **analysis using mathematical morphology: Part I**. IEEE Transactions on Pattern Analysis and Machine Intelligence 9, 4 (July 1987), 532–550.

HUANG, T. **Picture Processing and Digital Filtering**. Springer, New York. 1975.

JAIN, A.K. **Fundamentals of Digital Image Processing**, Prentice-Hall International 1989.

KERNIGHAN, BRIAN W. E RITCH, DENNIS M., C: **A Linguagem de Programação**. Rio de Janeiro, Campus, 1986.

KONAREV, V.G. et al.. Electrophoretic and serological methods in seed testing. **Seed science and technology**, Leningrad, 9: 807-17, 1981.

KOPF, JOHANNES; COHEN, MICHAEL; LISCHINSKI, DANI; UYTENDAELE, Matt. **Joint Bilateral Upsample**. SIGGRAPH 2007.

LOPES, C.R. – **Estudos sobre as relações filogenéticas entre algumas espécies de Coffea**. Botucatu, 1984. 169p.

MACHADO, L. P. B. **Descrição e análise de loci de microssatélites em populações naturais da espécie cactofílica Drosophila antonietae (DIPTERA; DROSOPHILIDAE)**. Tese de Doutorado. FMRP. Universidade de São Paulo, 104p., 2003.

MARTINEZ ERM, PAIVA LRS: **Eletroforese de ácidos nucléicos: uma prática para o ensino da genética**.

MATARREDONA, E. **O Uso do Processamento de Imagens Aplicadas na Radiologia Médica**, 1994.

MIZRAHI, VIVIANE V. **Treinamento em Linguagem C - módulo 1 e 2**, São Paulo, McGraw-Hill, 1990, 241p.

MORAES, R. M.. **Introdução aos Fundamentos Estatísticos do Processamento Digital de Imagens**. IX SINAPE - São Paulo - Brasil. 1990.

PARIS, SYLVAIN; KORNPROBST, PIERRE; TUMBLIN JACK; DURAND, FRÉDO. **A Gentle Introduction to Bilateral Filtering and its Applications**. ACM SIGGRAPH 2007.

PETSCHNINGG et al.. **Digital photography with flash and no-flash image pairs**. ACM SIGGRAPH conference (c) 2004, Association for Computing Machinery, Inc.

PLAUGER, P.J.; BRODIE J. STANDART. **C: guia de referência básica**, São Paulo, McGraw-Hill, 1991. 306p.

RABBANI M. **Digital Image Enhancement**. SPIE-Video Short Course Notes.1992.

ROSENFELD, A. **Image Analysis and Computer Vision – Computer Vision Graphics and Image Processing** – vol 59, p.367-404 – 1993.

RUBINSZTEIN, D.C.; AMOS, W.; LEGGO, J.; GOODBURN, S.; JAIN, S.; LI, S.; MARGOLIS, R.L.; ROSS, C.A.; FERGUSON-SMITH, M.A. **Microsatellite evolution –evidence for directionality and variation in rate between species**. Nature Genetics, v.10, p.337-343, 1995.

SCHILDT, HERBERT. **Turbo C: Guia do Usuário**, São Paulo, McGraw-Hill, 1988, 414p.

SERRA, J. **Image Analysis and Mathematical Morphology**, Volume 2: Theoretical Advances. Academic Press, New York, 1988.

SERRA, J. **Image Analysis and Mathematical Morphology**. Academic Press, London, 1982.

SHAW, DUNCAN J. **Electrophoresis**. London: Academic Press, c1969. 144 p.

SILVA, ANTÔNIO MACHADO E. **Curso Processamento digital de imagens de satélite**. Centro de Eventos da PUCRS - de 07 a 12 de outubro de 2001. Porto Alegre - RS. Disponível em [HTTP://www.cartografia.org.br](http://www.cartografia.org.br). Acesso em: 19 fev. 2007.

SOARES, L.F.S.; STEIN, L.H.; TIEPPO, E.; GEHLEN, M.; MORO, J.S.; PICHETH, G.; MARCHAUKOSKI, J.; HASS, I. E RAITTZ, R.T. **Análise Eletroforética em Géis Unidimensionais: Nova Abordagem Focada em Inteligência Artificial e Estudo Comparativo de Soluções**, Proc. VI Workshop de Visão Computacional, p. 66-71, Presidente Prudente, 2010.

SOILLE, P. **Morphological Image Analysis**. Springer Verlag, Berlin, 1999.

STERNBERG, S. Grayscale morphology. **Computer Vision, Graphics and Image Processing** 35 (1986), 333–355.

SUN, H.; KIRKPATRICK, B.W. **Exploiting dinucleotide microsatellite conserved among mammalian species**. Mammalian Genome, n.7, p.128-132, 1996.

TAUTZ, D. Hypervariability of simple sequence as a general source for polymorphic DNA markers. **Nucleic Acid Research**, v.17, n.16, p.6463-6471, 1989.

WILSON, GREGORY. **Programmer's Tool Chest – The OpenCV Library**. Disponível em <http://www.ddj.com/architect/184404319?pgno=1>. Acessado em 28/08/2011.

APÊNDICES

Imagens utilizadas neste trabalho.

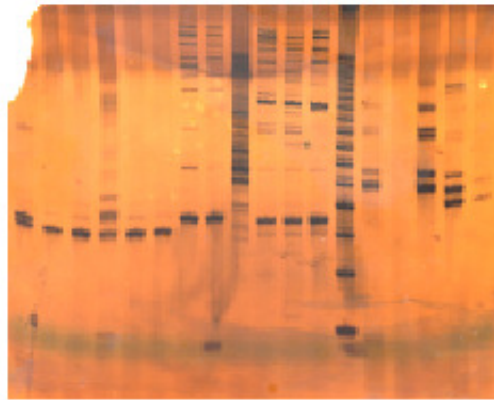


Imagem 1

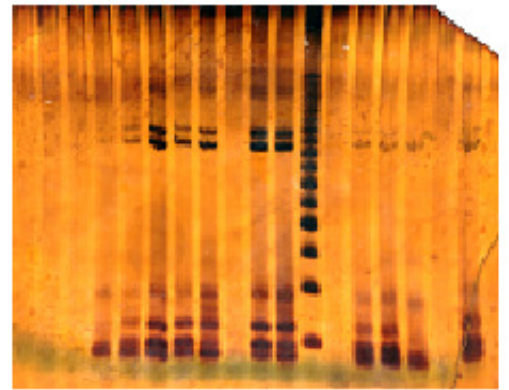


Imagem 2

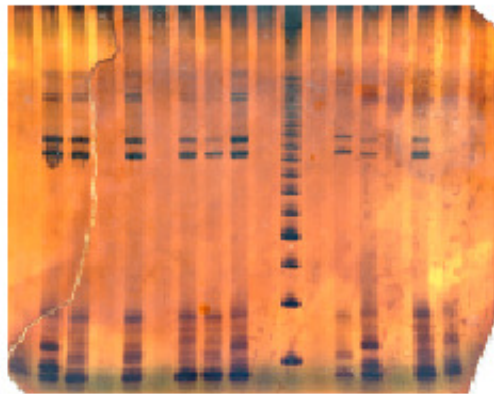


Imagem 3

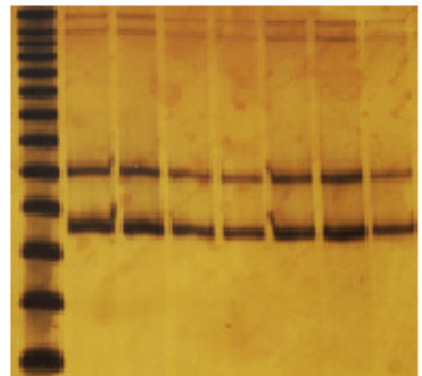


Imagem 4

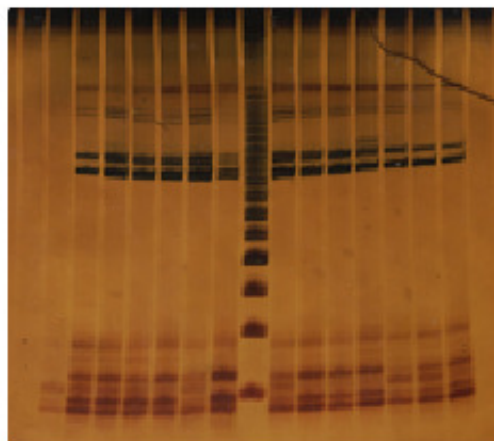


Imagem 5

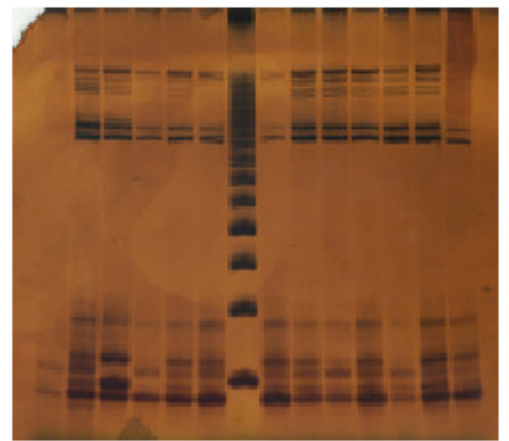


Imagem 6

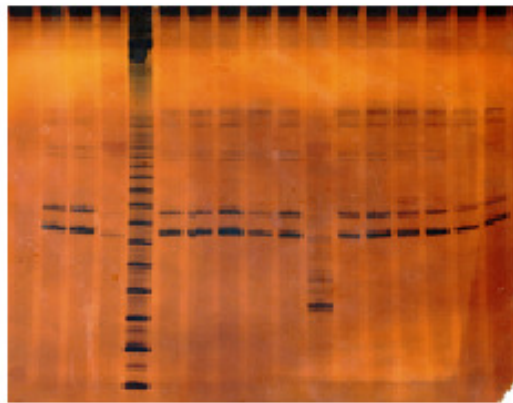


Imagem 7

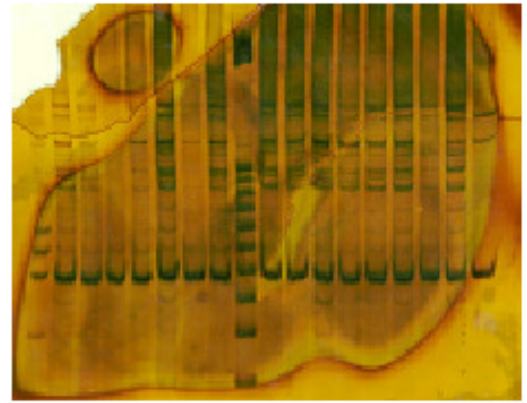


Imagem 8

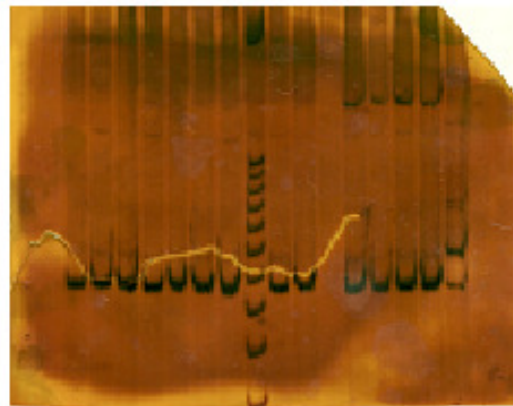


Imagem 9

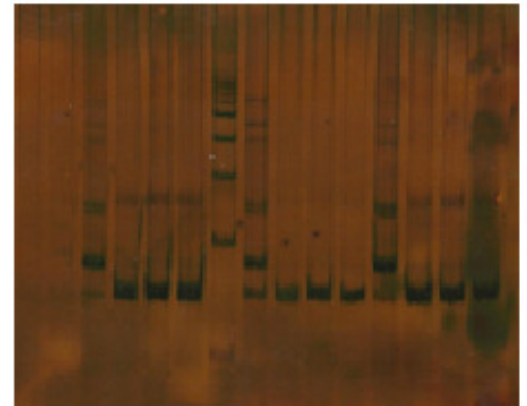


Imagem 10

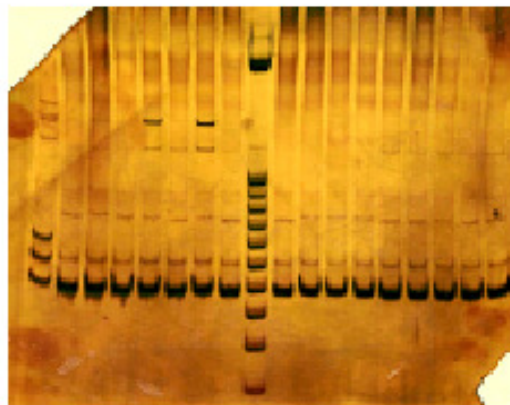


Imagem 11